Knee Osteoarthritis Severity Grade Binary and Multiclass Classification Using ROI-Based Diverse Features

Ravindra D. Kale[1], Dr. Sarika Khandelwal [2]

[1] Ph.D. Research Scholar, Computer Science & Engineering, G H Raisoni University, Amravati, India
[2] Associate Professor, Department of Computer Science & Engineering, G H Raisoni College of Engineering, Nagpur, India

ravindra.swati2012@gmail.com,sarikakhandelwal@gmail.com

Article Info	ABSTRACT
	Today, several diseaases have affected humans due to their unseen progression in early stages. Knee Osteoarthritis (OA) affecting most of the elderly mass is regarded as the common cause to restrict their normal activities and impairment. Therefore, a fast, speedy and cost effective early disgnosis of OA is the need of the hour. The work introduced in this article addresses the aforementioned issue and suggest an automated efficient diagnostic framework to classify OA severities based on X-ray or plain images. The Knee OA Severity Detetction Framework (K-OA-SDF) is used to study the impact on different class-based scenarios from binary to multiclass. The K-OA-SDF is subjected on X-ray images from Knee Osteoarthritis Dataset with Severity Grading (KODSG) of Kaggle dataset store to predict the knee OA grades. The K-OA-SDF is constructed using Region of Interest (ROI) Extraction Module, an efficient feature extraction module and a Convolutional Neural Network (CNN) based Classifier. The ROI module extract the significant portion of the X-ray images and enhances the contrast for beter quality features. The feature extraction module obtains diverse features from the ROI using image-based, object-based and traditional handcrafted features. Image-based features are acquired by dividing the ROI in two halves and averaging 16 columnar samples from both halves. Object-based blind features are obtained using modified VGG16 pre-trained network trained on the ImageNet dataset. Handcratfted features using quality descriptors are obtained to uplift the global and local details of the ROI. The CNN-based fine tuned classifier using 2000 (dimensionally reduced using Principal Component Analysis) features is able to outperform other state-of-the-art work in case of binary (98% accuracy) and ternay (88%) class categories while comparative results (71.17%) are obtained for five Kellegren Lawrence (KL) grades.

INTRODUCTION

Osteoarthritis (OA) is ranked the 11th most painful disease along with hip OA. This disability factor has affected people economically owing to the diagnosis cost and the treatment cost. It progression of the disease can be slowed down by systematically diagnosing the disease at the early stage which can significantly mitigate the impact of forthcoming disability and slow down its progression. Due to the absence of effective care for OA at the degraded stage besides surgery to replace the total joint, the only remedy available is its early diagnosis and behavioural interventions. This can help the patients to prolong their normal healthy life for certain years. The only clinical examination available at some specialized or private centers is Magnetic Resonance Imaging (MRI) which is somehow expensive and beyond the reach of people engaged in labour work or the mass below the poverty line. Also, MRI does not reflect the bone architecture which is significant for the OA progression.

Besides MRI, the safest and cheapest method used in clinical diagnosis for OA is widely available X-ray imaging. Despite the advantageous facts, X-ray imaging or Plain Radiography is almost insensitive to early changes in OA. OA is measured by the amount of wear caused to articular cartilage which is a tissue not reflected in an X-ray. A joint has 3D geometry and should be better evaluated using a 3D imaging technique whereas the X-ray uses only 2D sum projection. The most important aspect of knee OA is perfect evaluation which can be performed only by an experienced medical orthopaedic practitioner. Normally, the degeneration and wear of the cartilage are estimated by assessing the distance between two cartilage of the knee joint and changes in the bone structure. The fact that OA cannot be effectively diagnosed using X-ray and MRI imaging is costlier, making the clinical approach difficult. Today, we have no precise grading for the OA, and it depends on the subjectivity of the medical practitioner. Kellegren Lawrence (KL), a semi-quantitative grading scale that suffers from ambiguity due to expert disagreements. Thus, grade ambiguities in early diagnosis of OA impose a great challenge for clinicians in treating OA patients worldwide affecting millions of people.

Studies conducted in [1-3] concluded that merely using a single diagnosing technique is not sufficient and reliable for OA diagnosis. Therefore, Knee arthroscopy is a better approach to get a clear picture of the OA progression. The process involves inserting a tiny camera through small incisions to obtain a direct overview of the inside cartilage. The studies conducted also concluded that the non-invasive techniques (CT scans and MRI scans) are not so efficient and accurate enough due to their dependency on expertise. Till now, no specialized and specific diagnosing mechanism can accurately identify the degeneration and wear of the cartilage in the early stage of OA. Under such situations, the development of OA is highly influenced by certain biological and environmental factors [4]. Due to medical data privacy issues, not many samples with proper KL grades are publicly available for evaluation. OA is categorized as a big data problem and imposes a great challenge in identifying the accurate KL grade since different experts have different opinions on the grades. Therefore, there is an intense need for an automated diagnostic system for OA, especially using machine or deep learning techniques.

Today artificially intelligent systems have almost replaced and streamlined tasks that were performed manually earlier. Diagnosis of several diseases fully or partially automated using various machine learning techniques and deeply trained networks or through transfer-learned networks. The systems assisted medical experts in distinct applications for sound clinical analysis and decisions and relieved their burden by automating traditional manual routing experimentation [5]. Extensive use of DL models over benchmark datasets and transfer learning approaches have increased the popularity of DL networks in the field of biomedical engineering.

Work introduced in [6] used a bone finder tool to locate the patella and Linear Binary pattern (LBP) based texture features were provided to a CNN network for knee OA detection. The author used the X-ray images and evaluated the performance in terms of AUC and AP. Radiological features were extracted in [7] to analyze the performance of several ML tools along with the CNN and found that the CNN performed better as compared to ML techniques. Work in [8[was based on segmenting the knee area to improve in determining the minimum JSW. The JSW joint space was predicted for knee OA determination and their XGBoost classifier achieved 62.1% of AUC for KL grade and knee OA progression. Radiographs from a multicentre dataset were analyzed to predict the KL grades using a pre-trained residual network [9] and obtained 98% accuracy. Deep-learned features were extracted using two pre-trained models to classify 5-Kl grades in [10]. The authors used DenseNet201 and InceptionV3 for features and SVM to classify them and obtained 71.33% classification accuracy. A mean accuracy of 77.24% was obtained by [11] using a Deep-CNN model over X-ray images and classified early-stage OA severity grades. The work introduced in [12] detected knee edges for predicting the stages of knee OA using X-ray knee mages. The complex features thus obtained were then converted to simple feature vectors and ML techniques were used to predict the KL severity grades. A special attention module was used in [13] to extract features from the X-ray knee images while suppressing unnecessary information. Data from attention units were concatenated and a Mish's activation function was employed to improve convergence and model performance. Their attention unit-based model obtained an accuracy of 70.23%. An Otsu-based segmentation was applied to extract the region of interest from knee images in [14] and shape features from the ROI were used to classify the KL grades.

Although, several methods and materials have been used by researchers, and were able to obtain satisfactory results in categorizing the knee OA KL grades. Achieving promising accuracy for knee X-ray images in early-stage KL detection is still a goal for most researchers. This is because the early-stage knee OA images concerning the cartilage show a similar resemblance for different KL grades and available X-ray images are pruned to noise making it difficult to differentiate soft and hard tissues. Therefore, a robust feature extraction mechanism is required to finely differentiate the grade severities. Literature shows that merely considering only handcrafted features or features obtained through deep learning approaches is not sufficient to improve KL grade classification. Also, most of the existing approaches lack generalization ability when images from multiple datasets are considered.

The article based on K-OA-SDF contributes in the following aspects:

1.   An efficient generalized ROI extraction module is introduced to segment significant portion from the original X-ray knee images.
2.   Variety of features focussed directly on ROI image, blind features using modified VGG16, and coarse and fine handcrafted features are extracted to uplift global as well as local details from the ROI.
3.   The features are Max-normalized and reduced in optimal dimension using PCA to improve the CNN-based Classifier performance. The three stage K-OA-SDF performed better on KODSG kaggle dataset for two and multiclass categories realting to KL-grades.
The article is organized in the following respect: The next part of the paper deals with in-depth related work, section 3 deals with the description of the materials used along with the proposed framework, result and discussion is part of section 4, and the last section concludes the article.

RELATED WORK

The work introduced in [15] used two publicly available datasets containing 3026 and 4796 images. The data included belonged to both genders aged from 46 to 79 years old with five severity grades ranging from no OA to severe OA (0-4). The former dataset images from the MOST dataset were only used for training. The dataset was augmented by flipping the left knee and incorporating the ResNet34 pre-trained model for training. They rescaled the given region to a size of 300x300 and then obtained two square patches with a verticle offset. Patches were cropped from the lateral side and medial side which was then horizontally flipped. A two-branch Deep Siamese CNN model accepted two images to learn the symmetry measure between the two images. Each of the branches was constructed using a convolutional layer. A max pooling layer, a ReLu layer, a batch normalization layer, a global averaging layer, and a concatenation layer. They trained three separate models and selected the model which performed better on the validation data for each seed. Finally, they summed the predictions and subjected the sum to a softmax layer to evaluate the class probabilities of the OA KL grade. They concluded that their model for OA is suitable for fast diagnosis, and offers low cost. They obtained an average test accuracy of 66.71% for five KL grades. The authors in [16] carried out a six-stage preprocessing and augmentation of Knee images before feeding them to the Deep CNN for feature extraction followed by classification to 5-KL grades. In the pre-processing framework, they removed rows from the top and bottom of the image to extract the region of interest and make it a perfectly square image. Further, they resampled the image with cubic interpolation to obtain a 299x299 size image. Center pixel values were considered using the image mean and scaled pixel values using the standard deviation and normalized the image in the range [0 1]. The 1D image was replicated to obtain a 3D image and centered using the ImageNet RGB channel means which was followed by scale operation using standard deviation. They augmented the data using cropping, zooming, upscaling, noise addition, flipping, and adjusting contrast operations. They obtained 71% accuracy over the test set (4090 images) and compared the performance of only 50 images with evaluation from the best radiologists. The training set consists of 32116 images and features were revealed using a saliency map.
Deep Learning based on the integration of object detection by YOLO and visual transformer to improve the classification accuracy was introduced in [17]. The proposed framework uses an object detection YOLO-based CNN model to extract the bounding boxes of the knee from the input plain radiographs. The low-quality detection is

removed using the confidence scores (below 75%). In the preceding stage, spatial feature maps are obtained using a CNN from the ROI. The features thus obtained are flattened and used as a sequence to the successive stage. Finally, the visual transformer classifies the image to the respective KL grade by exploiting the correlation between different local textures. Their method performed better than others and classified five grades with 69.18% accuracy. They used 4796 (16-bit DICOM X-ray images) images from a single dataset which were collected from over 43100 imaging and clinical centers. Hybrid features using a combination of LBP and HOG with CNN classifiers were suggested in [18] to classify four KL grades of OA. The authors focused on extracting low-level features using the HOG descriptor and texture features using the LBP descriptor which were classified using three machine learning techniques. They used the Mendeley dataset IV [19] with 500 X-ray images of knee OA out of which 100 samples belonged to the normal or healthy category. The images were preprocessed by converting them to TIFF format to preserve the quality. Irrevelent details were removed during the conversion and opposite images were obtained for better ROI quality. The images were downscaled to reduce the computational cost and the mean value in 2x2 neighbourhoods was considered for evaluation. The images were then segmented to extract the region of interest and fed to the CNN and descriptors for feature extraction. Their multiclass classifier system was able to classify the samples with 97.14% accuracy for the four KL grades.

Two (binary and multiclass) independent frameworks for OA grade detection were introduced in [20]. In the first framework, a five-class classification model based on pre-trained CNN (feature extraction), PCA (dimension reduction), and SVM (classification) was used while in the second framework, the CNN model was fine-tuned to distinguish 2, 3, and 4 classes of OA. They found that for a low number of classes, the accuracy was better and degraded as the number of classes was increased from two to five. They evaluated their model performance over the Osteoarthritis Initiative dataset (OAI) (https://nda.nih.gov/oai/) and obtained an accuracy of 62% and 91% for 5 classes to 2 classes respectively. Augmenting the minority class and balancing the dataset for class imbalance was carried out in [21]. The sample images were rotated between -7 to +7 degrees and 90% samples were chosen for training the classifier. The test set comprised 45 samples from each class. ResNet34 was modified and two filly-connected layers were added to the network after an average pooling layer. The input shape was modified and a grayscale input was defined for the ResNet34 network. Two overlapping regions were extracted from the left and the right side of the ROI and fed to two separate identical modified ResNet34 networks. The right half region extracted was flipped before being input to the network. The final layer using the softmax activation function returned probabilities between 0 and 1 for each KL class. The proposed OA KL grade classification framework obtained 61.71% accuracy for a multiclass problem with Adam optimizer. Their model was unable to distinguish KL-0 and KL-1 classes.

A transfer learning approach using three different deep pre-trained networks was incorporated in [22]. The knee X-ray images were preprocessed involving filtering, resizing, and normalization operations. Features extracted from the ROI were then used for the classifier network for training and testing. The authors succeeded in obtaining an accuracy of 92.17% with the VGG16 network (other being CNN and ResNet50) for two classes viz. the normal and the affected class. Handcrafted features and blind features-based OA KL grade classification framework were presented in [23]. Handcrafted features including GLCM, Discrete wavelet transform, and LBP were fused to form 228 elements feature vector on one hand, and a Deep pre-trained network involving VGG19 and ResNet101 was employed to extract blind features separately with and without dimension reduction using PCA was subjected to a feed-forward neural network (FFNN) to classify knee images from RCU and OAI dataset in five KL grades. The framework obtained an average maximum accuracy of 99.1% and 98.2% on the OAI and RCU datasets respectively. The framework is computationally complex.

MATERIAL AND METHOD

This part of the paper explores the dataset description used for the experimentation and the methodology used to classify KL grades of OA. The Knee X-ray images are downloaded from one of the well-known publicly available high-quality datastore Kaggle [24] (Knee Osteoarthritis Dataset with Severity Grading (KODSG)). The dataset is

downloadable from (https://www.kaggle.com/datasets/shashwatwork/knee-osteoarthritis-dataset-with-severity) and can be easily accessed by researchers to develop and compare the performance of their KL-grade prediction frameworks. The dataset mainly consists of data for training, testing, and validation. Each of the folders further contains five subfolders consisting of images related to five KL grades labeled from 0 to 4. The KL-grades from 0 to 4 belong to subject knee images classified in five severity of osteoarthritis respectively named as healthy{0}(3218), doubtful{1}(1476),minimal{2}(2142),moderate{3}(1079), and severe{4}(249). The following table 1 shows the classification of the samples available in the dataset.

TABLE 1 - KNEE OA SAMPLES WITH KL-GRADE PRESENT IN THE DATASET.

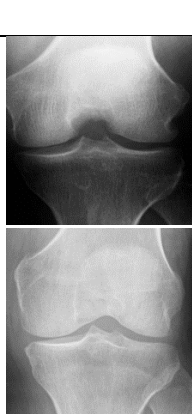| Description	KL-Grade	Number of Knee X-ray images			Total
		Training	Testing	Validation	
Healthy	0	2255	637	326	*3218*
Doubtful	1	1033	291	152	*1476*
Minimal	2	1491	441	210	*2142*
Moderate	3	753	222	104	*1079*
Severe	4	171	51	27	*249*
Total		*5703*	*1642*	*819*	**8164**

The total number of knee Osteoarthritis X-ray images as indicated in Table 1 are 8164 accommodating all the five severity grades. Except for KL grades 0 and 2, the samples corresponding to other KL grades are scarce and impose a great challenge for training and testing. Also, there is a data imbalance between the samples belonging to different classes. The severe class only has 249 samples with only 171 samples for training. The present approach does not consider improving the samples in the classes through augmentation but accepts the challenge by using the dataset in its original form. That is, the proposed K-OA-SDF framework uses only the available samples except that the validation and the test set are merged with the training set. Therefore table 1 was reduced to table 2 for the proposed K-OA-SDF framework. Further, 80% of the samples were used for training, and the rest of the 20% samples were used for testing from each of the KL grades. This way the training set was improved so that the classifier learns better and performs well in predicting the test sample class grade. Some of the sample X-ray images belonging to all the KL-grades are shown in Figure 1.

TABLE 2 - KNEE OA SAMPLES WITH KL GRADE CONSIDERED FOR THE PROPOSED WORK.

Description	KL-Grade	Total	Training	Testing
Healthy	0	*3218*	*2577*	*641*
Doubtful	1	*1476*	*1189*	*287*
Minimal	2	*2142*	*1714*	*428*
Moderate	3	*1079*	*852*	*227*
Severe	4	*249*	*199*	*50*
		8164	**6531**	**1633**

The original images from the dataset were preprocessed to extract the region of interest (ROI) and enhanced before multiple features were extracted to assist the classifier. The ROI extraction and enhancement were necessary due to three important aspects. Firstly, only the cartilage part, the upper and lower hard bone tissue, and the region in the vicinity of the hard tissues are significant part for determining the severity of OA. Secondly, the original knee images are not cleaned and the background tends to interfere with the foreground making it difficult to exactly identify the distance between the two hard tissues. Lastly, the unwanted lateral regions besides the knee joints

unnecessarily increase the dimension of the feature. Therefore, the upper and the lower regions of the knee, and unwanted lateral regions were eliminated. The effect of the background over the foreground was eliminated by subtracting the background from the foreground. Finally, the image was enhanced using the equalized histogram method. Figure 1 below shows some samples of knee X-ray images from all the severity classes (KL-grades). The intensity variations as seen from the sample images in figure 1 indicate that devising a common approach in the presence of such unevenness is difficult and poses a great challenge. Even the sample features belonging to the same class will be apart from each other to a great extent making the class overspread in the space. Due to this the nonlinearity between intra-class and inter-class samples is extensively high thus making the task of classification a complex issue.

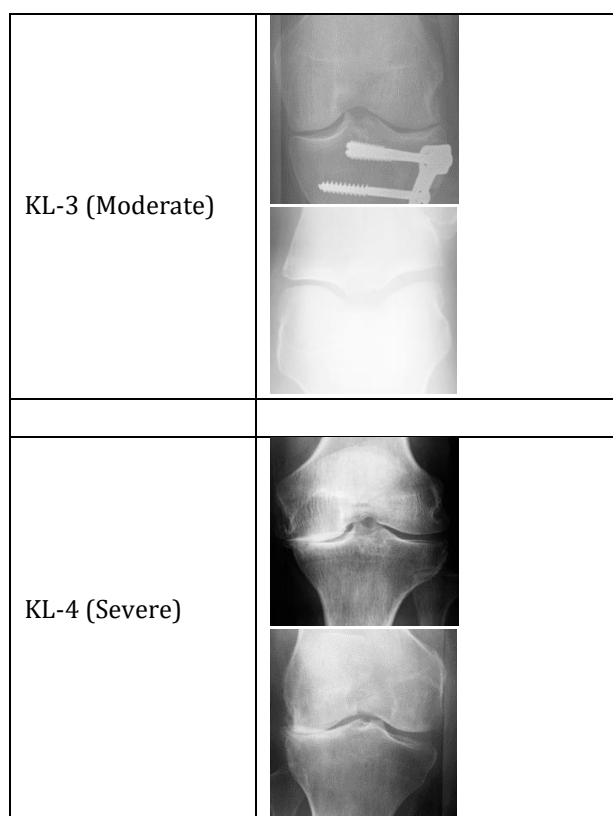| KL-Grade	X-ray Samples for the dataset
KL-0 (Healthy)	
KL-1 (Doubtful)	
KL-2 (Minimal)	

Figure 1 -Some of the sample X-ray images belonging to all the KL-grades

Thus the following Module shown in Figure 2 was used to segment the ROI from the original images and enhance the quality of the extracted ROI. The Image was Gaussian blurred using a 3x3 kernel and thresholded using a threshold value of 80. The threshold was determined through experimentation on several images from the dataset. This was done to find the actual knee portion (actual foreground) coverage in the image. The result of which can be seen in Figure 3.
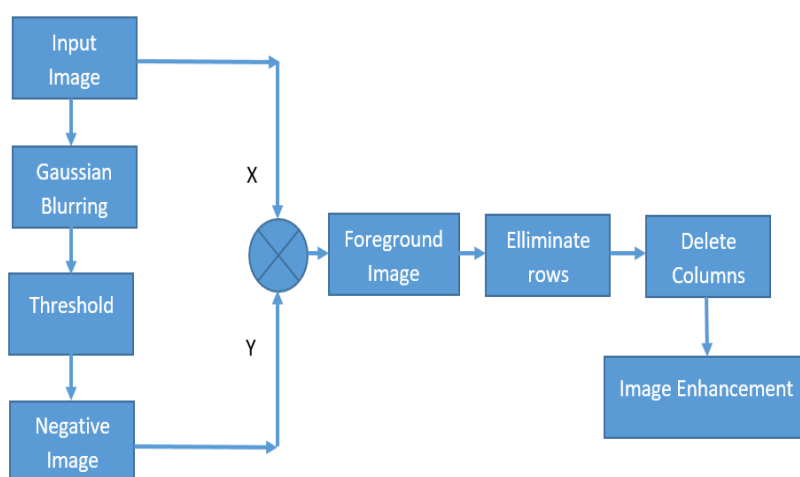

Figure 2 – The proposed ROI extraction and Enhancement Module

Figure 3– The ROI extraction process – Original image, Gaussian Blurred image (kernel – 3x3), and the Binary thresholded image (threshold – 80)

The binary image obtained is then inverted and multiplied with the original image to obtain the Masked image as shown below in Figure 4. The Masked image clearly shows the elimination of background from the foreground and locates the boundaries of the actual knee region. Also, the unwanted region beside the knee portion can be identified as compared to the original image where the edges of the knee region are blurred and cannot be identified.
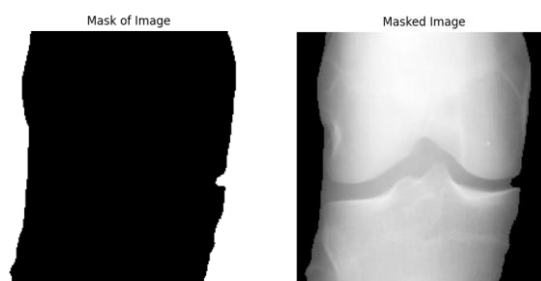

Figure 4 – The ROI extraction process – Foreground Mask, and the Masked image

To eliminate the upper and lower regions of the knee which have no significant contribution to OA, we tested all the images in the dataset with our experimentally found threshold value. We found that the vertical region of interest lies after 48 rows from the top and before 48 rows relating to the bottom. The upper threshold suited to 100% of samples from the dataset but using the same value for the lower threshold was a compromise. However, increasing the bottom threshold value above 48 was affecting some samples by cutting the cartilage portion while reducing the value below 48 caused an excessive amount of unwanted region to be added. Therefore, few images were cropped at the edge of the lower hard tissue connecting the cartilage but protected the actual region of the knee joint. Figure 5 shows the region of interest obtained after vertical crop operation. It also shows that there remains a comparable region laterally.

To eliminate the lateral unwanted part, the vertically cropped image was binarized, and the occurrence of the first non-zero along each column was found on both sides. For this, the binarized image was divided into two halves vertically. The first column containing a non-zero value on both sides was marked for cropping the image. The second figure of Figure 4 shows the result obtained using the mechanism. Experimental analysis of all the images in the dataset showed that few images bypassed the horizontal crop operation due to the knee region which extended to the image boundary. In such cases, there were no columns on either side with all zero elements.
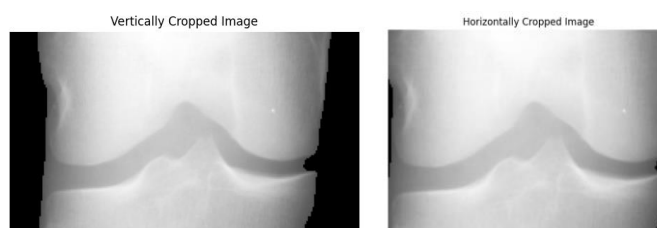

Figure 5 - Row cropped (along top and bottom) image, and column cropped (left and right) image

Different X-ray samples possessed different characteristics concerning their extreme boundaries, illumination, contrast, backgrounds, and foregrounds. As a result of which, the dimension of the extracted ROI varied with different images. For equal features, the ROI was resized to 128x128 after evaluating all the sample images in the dataset to avoid oversampling and undersampling which would eventually affect the details of the knee region. Figure 6 shows the result of the resizing operation where the details in the ROI are not clear enough to contribute the best feature descriptors. Therefore, to enhance the details in the ROI a contrast correction technique was adopted. The ROI was processed using a histogram equalization-based enhancement technique. The second figure of Figure 6 shows the difference between the actual ROI and the enhanced ROI image.
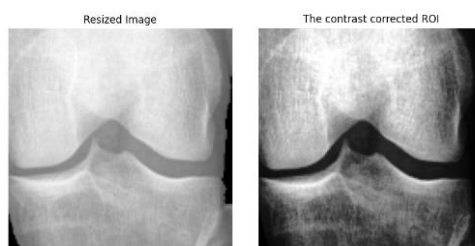

Figure 6 – Resized image (128 x 128), and the contrast-corrected image

KL-1 grade samples shown in figure 1 show exactly opposite images concerning the foreground and background relating to the pixel intensities. Such samples in the dataset are available in all classes. The proposed ROI extraction-enhancement mechanism is shown over one of the samples from KL-1 grade in Figure 7. This validates the generalization ability of the proposed ROI extraction in terms of positive and negative pixel intensities relating to foreground and background.

Figure 7 – Negative foreground-background sample images and the extraction and enhancement of ROI – Original image, cropped image, and the resized image

Highly illuminated samples such as the one shown in figure 1 for KL-3 grade were successfully segmented using the proposed ROI extraction technique. Figure 8 below shows the result of such a sample.

Figure 8 – Result of ROI extraction on highly illuminated knee sample – Original image, cropped image, and the resized image

The results obtained in Figures 6 to 8 indicate that the proposed ROI extraction and enhancement approach used in this work is efficient in segmenting the required region of interest irrespective of the illumination conditions of the X-ray knee images. The next stage deals with the feature extraction from the ROI image. Many researchers have

concluded that merely using handcrafted features or blind features does not improve the performance of a classifier. The resemblance between different sets of classes is high, especially between normal and doubtful KL grades, doubtful and moderate classes, and minimal and severe classes. Therefore, the feature descriptors should not only collect the global details but also fine disparities between different classes so that feature elements subject very low burden on the classifier. The proposed feature extraction Module shown in Figure 9 is used to extract distinct features from the ROI image.
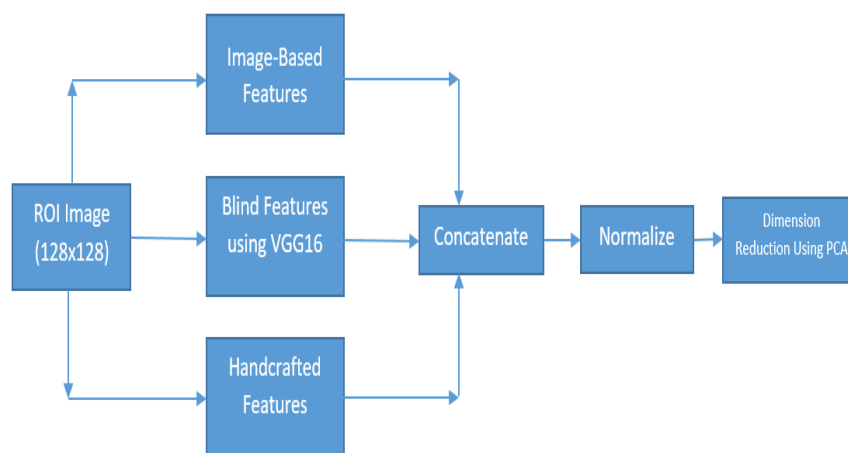

Figure 9 – The proposed Feature Extraction Module

Image-based features were directly obtained by dividing the image into two halves, the left and the right half. This was done to capture the groove (distance between the upper and the lower hard bone tissues) that lies between the upper and the lower hard tissues on both sides. A mean of 16 columns was computed and all the 128x1 vectors were vertically concatenated to obtain a feature vector of 1024x1 for each of the halves. Finally, the 1024x1 feature vector belonging to both halves was concatenated and added to the feature set.

VGG16 pre-trained network which is trained using the 'ImageNet' dataset consisting of a large number of distinct images can extract blind features from an input image. It has been used in many applications by modifying the output layers. The final layer was removed and the prior input was flattened. Two fully connected 'Dense' layers were added to the existing network of dimensions 2048 followed by 1024. The layers were activated using the '*relu*' activation function. The 2D ROI-enhanced image was converted to 3D by replicating the 2D frame along the second and third dimensions. The 3D image was then preprocessed (values formatted in a specific range) and applied as input to the modified VGG16 network. The output of VGG16 a 1x1024 feature vector representing the image blind features was sampled at a low rate of 8 samples by considering the average. The samples were averaged to eliminate the zero values from the 1024 elements.

A variety of handcrafted features were obtained to uplift global and local details along with textural and edge information from the Knee X-ray image. Patch-based textural information was acquired using a 3x3 neighborhood of the 128x128 image. The textural details were then averaged over a 5x5 block or window to reduce the dimension of the feature set.

Edge-based features were obtained using a 3x3 kernel whose center element value=8 while the neighbors have value=-1. A 2D filter using the kernel was used over the 128x128 dimension grayscale image and the coefficients of the filter were summed along both axis (axis=0 and axis=1) and concatenated to obtain a 128+128=256 element vector thus detailing the edge information from the input image. Another edge informative filter using the 'Sobel' operator was used with the value of sigma = [0.5, 1, 1.5, 2], length of the filter L = [9, 11, 13, 15] (length in Y-direction), and 12 orientations from [0 to 165 at an offset of 15) to obtain 32 elements in the features. Using two such edge operators, minimum loss due to edge miss was ensured.

Wavelet-based features using six different mother wavelets ['bior3.1', 'bior3.5', 'bior3.7', 'db3', 'sym3', 'haar'] were obtained from the vertical and diagonal coefficients in terms of magnitude and energy. The absolute and the square

values of the vertical coefficients and the diagonal coefficients are summed up separately to measure the magnitude and energy of both components. These two measures provide the information concentration in both directions. The effect due to the horizontal component was ignored since the measure was nearer to the vertical details. Using six mother wavelets with two measures resulted in 24 values contributing to the feature set.

The overall image contrast, energy, homogeneity, correlation, ASM, and dissimilarity were measured using the Gray Level Co-occurrence Matrix. Global-level indicators concerning 6 parameters were used to enhance the performance of the classifier. Fine details from the grayscale image were extracted using an LBP descriptor using a radius of 3. The texture pattern resulting from the LBP descriptor was summed along rows and columns and then concatenated to form a 256-element feature vector. Lastly, the complete image was divided into 4x4 non-overlapping blocks and the mean was computed to obtain a 32 x 32=1024 element feature vector. The summary of the feature set is listed in Table 3 with their respective parameters and dimensions.

TABLE 3 – VARIOUS DESCRIPTORS USED FOR HANDCRAFTED FEATURES WITH PARAMETERS AND DIMENSION

Descriptors	Parameters	Dimension
LBP Fine	3x3 patch Averaging over 5x5 window	1024
Matched Filter (Edge-Based)	3x3 Kernel (Neighbourhood value = -1 and center pixel = 8)	256
Matched Filter (Sobel-Based)	Sigma = [0.5, 1, 1.5, 2] L = [9, 11, 13, 15] Orientation - 12	32
Wavelet (Six Mother wavelets)	['bior3.1', 'bior3.5', 'bior3.7', 'db3', 'sym3', 'haar'] Magnitude and Energy of Vertical and Diagonal components	24
GLCM	Parameters = 6	6
LBP	Radius = 3	256
Original Image	Mean over 4x4 patch	1024
Total Feature Vector Length		**2622**

Table 4 below shows the total number of features that were obtained from all three branches of Figure 8. These are image-based features, blind features using the VGG16 pre-trained network, and handcrafted features. The total number of features that represented the ROI image was 4798.

Table 4 – Summary of features set

Sr. No.	Name of the Feature	Number of Features
1	Image-Based	2048
2	Blind	128
3	Handcrafted or Traditional	2622
Features per ROI image		**4798**

Since the features were obtained using different descriptors, individual features were normalized using the Max-Normalization technique. That is, features corresponding to all the 8164 images were computed and placed along the rows of an array. The columns of the array were normalized by finding the maximum along the columns and dividing each column element by the maximum value. Further, the normalized feature vector was dimensionally

reduced using the PCA and resting the feature-length for an ROI image to 2000. All the 2000 components were used for classification and the value of 2000 was found experimentally by checking the classifier performance.

RESULTS AND DISCUSSION

Experiments were conducted on the five KL-grade samples to determine the correlation between them. We considered 9 different combinations between the KL grades and distributed them in a different set of classes. The five classes or grades were resolved in 2, 3,,4, and finally, five categories to evaluate the performance of the proposed preprocessing+feature extraction Module. By categorizing the different KL-grade samples into different numbers of classes, we obtained a clear picture of the similarity of the samples between different Kl-grades. To achieve the objective, two or more KL-grade samples were grouped into a single class for a particular combination, and the same were put in different classes for another combination. The following Table 5 shows the classification accuracy obtained for different sets of combinations. We evaluated the K-OA-SDF framework on two, three, four, and five classes.

The feature set for 8164 images obtained after normalization followed by dimension reduction was separated into training and test sets. 80% of random samples were selected for training and the remaining for test purposes from all the two, three, four, and five categories. We constructed a sequential network with three 1D convolutional layers and a Dense layer, the kernel size being 3 for each of the 1D convolutional layers. All three convolutional layers were activated using the 'selu' function and the first layer was followed by a DropOut layer with a factor of 0.1. The classification was achieved by employing a Softmax function in the Dense layer. The batch size was set to 15 samples and iterated for 10 epochs. The number of neurons used with the convolutional layers was 1000, 150, and 150 respectively. The following Figure 10 shows the network architecture used for classifying the severity grades of the Knee OA.
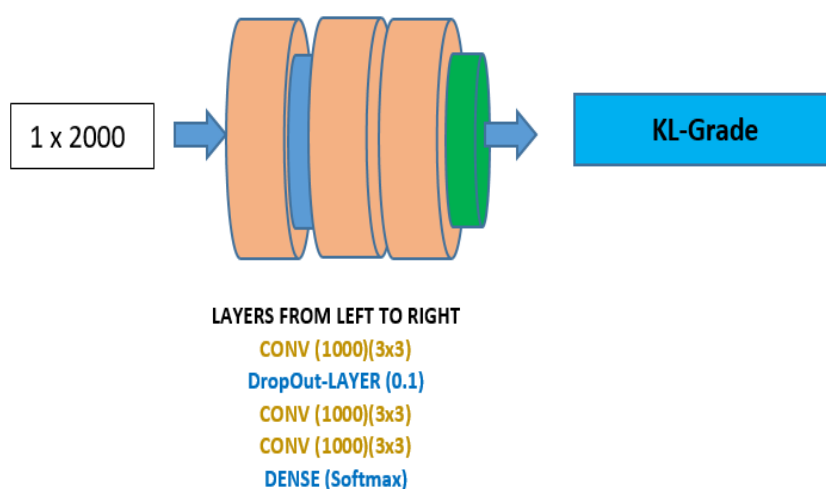

LAYERS FROM LEFT TO RIGHT
CONV (1000)(3x3)
DropOut-LAYER (0.1)
CONV (1000)(3x3)
CONV (1000)(3x3)
DENSE (Softmax)

Figure 10 – The CNN Classifer to distinguish KL-Grades

A simple visual look at different KL-grade samples reveals that the doubtful samples (KL-1) show higher similarity to the normal samples (KL-0). Also some of the samples belonging to Kl-2 and KL-3 are ambiguous pertaining to their present classes due to absence of significant disparities between them. The moderate and the severe class however are distinguishable due to significant bone structures. The upper and the lower hard tissues shows no cartilage between them in the latter case and can be detected by quality descriptors which can learn better at the knee joint. For accurate diagnosis of knee OA from X-ray images, expertise and experience of medical prationer is important. However, the accuracy of diagnosis remains a question and can be only solved using an automated system to classify the patients in proper grades. Based on the facts represented by different researchers, we initially tried to analyze the resemblances of different classes by categorizing the Kl-grades.

When low and high severity OA KL-grades were partitioned, the classification accuracy was higher. For two classes, Cx=1, the accuracy was 98.28%. As the severity grade KL-4 is extended towards the moderate, minimal and doubtful region, the classification accuracy shows declination from 91.18% to 82.19. When the samples were distributed in three categories as in the case of Cx=5 to Cx=7, the average accuracy was found to be approximately 74%. We considered only one combination for four class problem since there is high similarity between the KL-1 and KL-2 with respect to other grades. In last two cases, we fine tuned the network hyperparameters and the best accuracy found were 78% and 71.17% respectively.

TABLE 5 – EXPERIMENTAL ANALYSIS FOR VARIOUS KL-GRADE COMBINATIONS

Cx	KL-0	KL-1	KL-2	KL-3	KL-4	% Accuracy	Description
1	0				1	**98.28**	Two-Class
2	0			1		91.18	
3	0		1			86.09	
4	0	1				82.19	
5	0		1		2	**87.99**	Three-Class
6	0		1	2		75.22	
7	0	1		2		74.25	
8	0	1	2	3		**78**	Four-Class
9	0	1	2	3	4	**71.17**	Five-Class

We compared the performance of our proposed Knee OA classification framework with other state-of-the-art techniques found in the literature. As shown in Table 5, we have considered only one combination for 4-classes while other combinations of KL-grades are also possible. Table 6 shows the comparison based on different datasets (mostly the MOST (Multicenter Osteoarthritis Study - https://most.ucsf.edu/multicenter-osteoarthritis-study-most-public-data-sharing) and the OAI (Osteoarthritis Initiative - https://nda.nih.gov/oai) dataset) for different categoies relating to the combination of KL-grade samples. For 2 and 3 classes, the proposed K-OA-SDF framework outperformed others. The Deep Hybrid Learning Model II suggested in [20] showed superior performance as compared to others in case of 4 and 5 classes using the OAI dataset. For the 4-class scenario, other combinations were not evaluated, however we obtained 78% accuracy. Also, the number of samples available with the OAI and the MOST datasets are more as compared to the KODSG dataset which improves the learning of the classifier. In similar manner, DHL-I method suggested in [20] achieved a higher accuracy with the OAI dataset of 74.57% while we obtained 71.17% with the KODSG dataset. Using a simple network, better results are obtained especially for the 2-class and the 3-class case. The perfromnace can be improved by using transfer learning with pre-trained networks.

TABLE 6 – PERFORMANCE COMPARISON FOR FIVE KL-GRADE CLASSIFICATION

Reference	Method	Year	Dataset	Classes	Accuracy
[20]	DHL-II	2022	OAI	2	90.8
[29]	ML	2019	OAI	2	82.98
[31]	DNN	2020	OAI	2	79.6
[33]	ResNet101	2023	OAI	2	89
Proposed Work		**2024**	**KODSG**	**2**	**98.28**

[8]	DCNN	2021	MOST+OAI	3	73.46
[20]	DHL-II	2022	OAI	3	87
[33]	ResNet101	2023	OAI	3	83
Proposed Work		**2024**	**KODSG**	**3**	**87.99**
[26]	DenseNet Ensemble	2018	OAI	4	78.37
[20]	**DHL-II**	**2022**	**OAI**	**4**	**88**
Proposed Work		2024	KODSG	4	78
[25]	CNN	2017	MOST+OAI	5	61.9
[27]	Siamese	2018	MOST+OAI	5	66.71
[28]	VGG19	2019	OAI	5	69.70
[30]	ResNet169	2020	OAI	5	70.66
[8]	DCNN	2021	MOST+OAI	5	66.68
[32]	CNN + Attention	2021	OAI	5	69.18
[20]	CNN	2022	OAI	5	62
[20]	**DHL-I**	**2022**	**OAI**	**5**	**74.57**
[33]	ResNet101	2023	OAI	5	69
Proposed Work		**2024**	**KODSG**	**5**	**71.17**

CONCLUSION

A sophisticated knee OA severity detection framework (K-OA-SDF)is presented in this article which includes ROI extraction mechanism, ROI enhancement, diverse feature extraction model and the CNN based classifier. The preprocessing stage of the framework posses the generalization ability to segment all the X-ray knee images from the dataset without losing the significant part. It rules out the chances of ignoring samples from the dataset. This had been explained with respect to Figure 7 and Figure 8 where there is abrupt change concerning the illuminations. The efficient feature extraction module concentrates on image level features, object based features using the VGG16 network and variety of local and global handcrafted features to uplift overall as well as fine details from the ROI. Most of the features were reduced in dimension by averaging over certain samples to reduce the burden of the classifier and eliminate zero values. The K-OA-SDF outperform other state-of-the-art works for binary and ternary classes. The performance of the proposed K-OA-SDF over all the five KL-grades is comparative to the best technique DHL-II [20]. The performance can be improved by eliminating uneven samples from the dataset which possess negative illuminations from other normal X-ray images. Pre-trained models using transfer learning with finely tuned hyperparameters can enhance the overall performance.

The top fully connected layer of the VGG16 network can be modified for better performance. Better discriminative features such as the knee joint distance would have proven fruitful. Small variation between KL-grades 0 to 2 requires powerful descriptors to address and similarly for last two grades, the moderate and the severe. The crucial concern realted to the dataset is its class imbalance. Some of the classes have been poorly represented while others are significant. Consequently, the problem imposes great challenge on the framework and increases the risk in detecting the minority class properly.

Despite all these issues and limitations, our K-OA-SDF puts a clear picture about the correlation of samples between different classes. To conclude, the K-OA-SDF introduced in this article contributes to early detection of OA, help to mitigate the progression of OA, and improve humans quality of life.

Declaration of Interests:- Authors have no conflicts of interest.

Data Availibility :- Dataset is pubiicaly available and can be downloaded from kaggle.com.

REFERENCES

[1] Donders, C.M.; Spaans, A.J.; Bessems, J.H.; van Bergen, C.J. Arthrocentesis, arthroscopy or arthrotomy for septic knee arthritis in children: A systematic review. J. Child. Orthop. 2021, 15, 48–54.

[2] Radhamony, N.G.; Walkay, S.; Palaneer, S.; Hamadto, M. Predictors of failure after initial arthroscopic washout in septic arthritis of native knee joint-a retrospective analysis. Ann. Med. Surg. 2022, 74, 103269.

[3] Jaffe, D.; Costales, T.; Jauregui, J.J.; Koenig, S.; Weir, T.B.; Greenwell, P.; Christian, M.; Henn, R.F., III. Current surgical practice for septic arthritis of the knee in the United States. J. Orthop. 2021, 25, 88–92.

[4] Kokkotis, C.; Moustakidis, S.; Papageorgiou, E.; Giakas, G.; Tsaopoulos, D.E. Machine learning in knee osteoarthritis: A review. Osteoarthr. Cartil. Open 2020, 2, 100069.

[5] Roy, S.; Meena, T.; Lim, S.-J. Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine. Diagnostics 2022, 12, 2549.

[6] Bayramoglu, N.; Nieminen, M.T.; Saarakkala, S. Machine learning-based texture analysis of patella from X-rays for detecting patellofemoral osteoarthritis. Int. J. Med. Inform. 2022, 157, 104627.

[7] Cheung, J.C.W.; Tam, A.Y.C.; Chan, L.C.; Chan, P.K.; Wen, C. Superiority of multiple-joint space width over minimum-joint space width approach in the machine learning for radiographic severity and knee osteoarthritis progression. Biology 2021, 10, 1107.

[8] Tiulpin, A.; Saarakkala, S. Automatic grading of individual knee osteoarthritis features in plain radiographs using deep convolutional neural networks. Diagnostics 2020, 10, 932.

[9] Javed Awan, M.; Mohd Rahim, M.S.; Salim, N.; Mohammed, M.A.; Garcia-Zapirain, B.; Abdulkareem, K.H. Efficient detection of knee anterior cruciate ligament from magnetic resonance imaging using deep learning approach. Diagnostics 2021, 11, 105.

[10] Teo, J.C.; Khairuddin, I.M.; Razman, M.A.M.; Majeed, A.P.A.; Isa, W.H.M. Automated Detection of Knee Cartilage Region in X-ray Image. Mekatronika 2022, 4, 104–109.

[11] Tri Wahyuningrum, R.; Yasid, A.; Jacob Verkerke, G. Deep Neural Networks for Automatic Classification of Knee Osteoarthritis Severity Based on X-ray Images. In Proceedings of the 8th International Conference on Information Technology: IoT and Smart City, Xi'an China, 25–27 December 2020; ACM International Conference Proceeding Series. Volume Part F168341, pp. 110–114.

[12] Xiao, Y. Using Machine Learning Tools to Predict the Severity of Osteoarthritis Based on Knee X-ray Data. Ph.D. Thesis, Marquette University, Milwaukee, WI, USA, 2020.

[13] Feng, Y.; Liu, J.; Zhang, H.; Qiu, D. Automated grading of knee osteoarthritis X-ray images based on attention mechanism. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1927–1932.

[14] Chan, S.; Dittakan, K.; El Salhi, S. Osteoarthritis detection by applying quadtree analysis to human joint knee X-ray imagery. Int. J. Comput. Appl. 2022, 44, 571–578.

[15] Aleksei Tiulpin, Jerome Thevenot, Esa Rahtu, Petri Lehenkari, and Simo Saarakkala, "Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach," Scientific Reports, 2018, 8:1727.

[16] Kevin A. Thomas, Lukasz Kidzinski, Eni Halilaj, Scott L. Fleming, Guban R. Venkataraman, Edvin H. G., Garry E. Gold, and Scott L. Delp, "Automated Classification of Radiographic Knee Osteoarthritis Everity Using Deep Neural Networks," Radiology: Artificial Intelligence, 2020, 2(2): e190065.

[17] Yifan Wang, Xianan Wang, Tianning Gao, Le Du, and Wei Liu, "An Automatic Knee Osteoarthritis Diagnosis Method Based on Deep Learnings: Data from the Osteoarthritis Initiative," Journal of Healthcare Engineering, Volume 2021, Article ID 5586529.

[18] Mahum, R.; Rehman, S.U.; Meraj, T.; Rauf, H.T.; Irtaza, A.; El-Sherbeeny, A.M.; El-Meligy, M.A. A Novel Hybrid Approach Based on Deep CNN Features to Detect Knee Osteoarthritis. Sensors 2021, 21, 6189.

[19] Gornale, S.S.; Patravali, P.U.; Hiremath, P.S. Automatic Detection and Classification of Knee Osteoarthritis Using Hu's Invariant Moments. Front. Robot. AI 2020, 7, 591827.

[20] Ahmed, S.M.; Mstafa, R.J. Identifying Severity Grading of Knee Osteoarthritis from X-ray Images Using an Efficient Mixture of Deep Learning and Machine Learning Models. Diagnostics 2022, 12, 2939.

[21] Cueva, J.H.; Castillo, D.; Espinós-Morató, H.; Durán, D.; Díaz, P.; Lakshminarayanan, V. Detection and Classification of Knee Osteoarthritis. Diagnostics 2022, 12, 2362.

[22] Alshamrani, H.A.; Rashid, M.; Alshamrani, S.S.; Alshehri, A.H.D. Osteo-NeT: An Automated System for Predicting Knee Osteoarthritis from X-ray Images Using Transfer-Learning-Based Neural Networks Approach. Healthcare 2023, 11, 1206.

[23] Khalid, A.; Senan, E.M.; Al-Wagih, K.; Ali Al-Azzam, M.M.; Alkhraisha, Z.M. Hybrid Techniques of X-ray Analysis to Predict Knee Osteoarthritis Grades Based on Fusion Features of CNN and Handcrafted. Diagnostics 2023, 13, 1609.

[24] Chen, Pingjun (2018), "Knee Osteoarthritis Severity Grading Dataset", Mendeley Data, V1, doi: 10.17632/56rmx5bjcr.1.

[25] Antony, J.; McGuinness, K.; Moran, K.; O'Connor, N.E. Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity Using Convolutional Neural Networks. In International Conference on Machine Learning and Data Mining in Pattern Recognition; Springer: Berlin/Heidelberg, Germany, 2017; pp. 376–390.

[26] Norman, B.; Pedoia, V.; Noworolski, A.; Link, T.M.; Majumdar, S. Applying Densely Connected Convolutional Neural Networks for Staging Osteoarthritis Severity from Plain Radiographs. J. Digit. Imaging 2019, 32, 471–477.

[27] Tiulpin, A.; Thevenot, J.; Rahtu, E.; Lehenkari, P.; Saarakkala, S. Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. Sci. Rep. 2018, 8, 1727.

[28] Chen, P.; Gao, L.; Shi, X.; Allen, K.; Yang, L. Fully Automatic Knee Osteoarthritis Severity Grading Using Deep Neural Networks with a Novel Ordinal Loss. Comput. Med. Imaging Graph. 2019, 75, 84–92.

[29] Brahim, A.; Jennane, R.; Riad, R.; Janvier, T.; Khedher, L.; Toumi, H.; Lespessailles, E. A Decision Support Tool for Early Detection of Knee OsteoArthritis Using X-Ray Imaging and Machine Learning: Data from the OsteoArthritis Initiative. Comput. Med. Imaging Graph. 2019, 73, 11–18.

[30] Thomas, K.A.; Kidzi ́nski, Ł.; Halilaj, E.; Fleming, S.L.; Venkataraman, G.R.; Oei, E.H.G.; Gold, G.E.; Delp, S.L. Automated Classification of Radiographic Knee Osteoarthritis Severity Using Deep Neural Networks. Radiol. Artif. Intell. 2020, 2, e190065.

[31] Moustakidis, S.; Papandrianos, N.I.; Christodolou, E.; Papageorgiou, E.; Tsaopoulos, D. Dense Neural Networks in Knee Osteoarthritis Classification: A Study on Accuracy and Fairness. Neural Comput. Appl. 2020, 5, 1–13.

[32] Wang, Y.; Wang, X.; Gao, T.; Du, L.; Liu, W. An Automatic Knee Osteoarthritis Diagnosis Method Based on Deep Learning: Data from the Osteoarthritis Initiative. J. Healthc. Eng. 2021, 2021, 5586529.

[33] Mohammed, A.S.; Hasanaath, A.A.; Latif, G.; Bashar, A. Knee Osteoarthritis Detection and Severity Classification Using Residual Neural Networks on Preprocessed X-ray Images. Diagnostics 2023, 13, 1380.