

Received 8 February 2023, accepted 13 February 2023, date of publication 22 February 2023, date of current version 8 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3247502



RESEARCH ARTICLE

Knee Osteoarthritis Detection Using an Improved CenterNet With Pixel-Wise Voting Scheme

SULIMAN ALADHADH^{ID1} AND RABBIA MAHUM^{ID2}

¹Department of Information Technology, College of Computer, Qassim University, Buraydah 52571, Saudi Arabia

²Department of Computer Science, UET Taxila, Taxila 47050, Pakistan

Corresponding author: Rabbia Mahum (rabbia.mahum@uettaxila.edu.pk)

This work was supported by the Deanship of Scientific Research, Qassim University.

ABSTRACT To detect knee disease, radiologists have been utilizing multi-view images such as computer tomography (CT) scans, MRIs, and X-rays. The cheapest method is X-ray to attain the images that is used widely. There exist various image processing techniques to detect knee disease at the initial stages; however, there is still room for improved accuracy and precision of the existing algorithms. Furthermore, in machine learning-based techniques, hand-crafted feature extraction mechanism is a tedious task. Therefore, in this paper, we suggest a technique based on customized CenterNet with a pixel-wise voting scheme to extract the features automatically. Our model uses the most representative features due to the best localization results and a weighted pixel-wise voting scheme which takes input from a predicted bounding box using modified CenterNet. It gives a more accurate bounding box based on the voting score from each pixel inside the former box. Moreover, we employed the distillation knowledge concept to make our model simple without increasing its computational cost, and transfer knowledge from a complex network to a simple network. Therefore, our proposed model detects the KOA in knee images precisely and determines its severity level according to the KL grading system such as Grade-I, Grade-II, Grade-III, and Grade-IV. Our proposed model is a robust and improved architecture based on CenterNet utilizing a simple DenseNet-201 as a base network for feature extraction. Due to the dense blocks employed in a base network, most representative features are extracted from the knee samples. We employed two benchmarks i.e. Mendeley VI for the training, and testing, and the OAI dataset for cross-validation. We evaluated the performance of the proposed technique using various experiments and it is estimated that our proposed model outperforms the existing techniques with an accuracy of 99.14% over testing and 98.97% over cross-validation.

INDEX TERMS Machine learning, detection performance, HCI, classification, deep learning, multi-scale features.

I. INTRODUCTION

Knee Osteoarthritis (KOA) is a chronic joint disease due to the worsening of articular cartilage in the knee. The symptoms of KOA comprise joint noises due to cracking, swelling, pain, and difficulty in movement. Moreover, the severe symptoms of KOA may cause fall incidents i.e. fracture in the knee bone that ultimately results in disability of leg [1]. Various imaging techniques which have been employed for the analysis of knee disease include MRI, X-ray, and CT scans. Furthermore, MRI and CT scans are also considered suitable

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan^{ID}.

for KOA assessment [2], [3]. They are accompanied using an intravenous contrast agent [4] which provides a clear view of Knee joints. However, these approaches are associated with high costs, increased examination time, and potential health risks such as patients with renal inadequacy [5]. Therefore, there should be some techniques for the assessment of KOA that can be employed without the contrast agent and require minimum expense, and time of examination. Therefore, an X-ray is considered a more feasible way to provide bony structure visualization and is a less expensive approach for knee analysis.

Cartilage helps in flexible movement, however, when it decreases with age or any accidental loss, it causes the disease

Knee Osteoarthritis (KOA). The Knee joint is composed of two bones i.e. tibia and femur. Both of these bones are joined with the thick material that is called cartilage. The severity of the disease is measured through a grading system known as Kellgren & Lawrence (KL) which is based on the radiographic classification of KOA. It comprises of 4 grades i.e. Grade I, Grade II, Grade III, and Grade IV [6]. Grade I shows the lowest severity of the disease whereas Grade IV refers to the highest severity level. Early detection and the classification of disease help physicians to treat patients with a high success rate. The most common cause of KOA is being overweight and the disease progresses towards higher grade with the age. Moreover, the average age of forty-five years of KOA patients has been reported in [7]. In the USA, patients of KOA having an age of 65 years or above have been assessed for KOA through radiography [6] and more than twenty-one million people have this disease [8]. In Asian countries, this disease has been spreading day by day. In Pakistan, 25% of the rural area and 28% of the urban population has KOA disease [9]. Besides medication, KOA can be treated through exercise, weight reduction, walking, and physiotherapy [10]. There exist various techniques for KOA detection and classification such as Gait Analysis, MRI, Impedance Signals, etc. [11], [12]. Knee joint width space is an important key factor to assess the KOA severity. Therefore, X-rays help in the visualization of joint width space and MRI assesses the cartilage thickness and complete surface condition. On the other side, bioelectric impedance signals are the most useful approach for detecting the KOA. It requires a low expense and is easy to employ [13].

There exist various ML and DL-based methods for the detection and classification of KOA [10], [14], [15], [16], [17], [18]. In [19], a model has been developed for KOA detection and classification based on the hybrid feature descriptors such as HOG and CNN employed with the KNN clustering algorithm. The algorithm outperformed the existing techniques, attaining an accuracy of 97.14%. However, in this study, we aim to develop a system based on deep learning that has low complexity and gives better accuracy for all grades of KOA rendering to the KL grading system.

The advancement in segmentation based techniques has also gain importance in last two decades. The images pixels are pictorial elements to discern the various regions of the input samples. Segmentation is a technique to divide the whole image into various regions based on the application requirement [20], [21], [22]. These segmentation-based techniques play a vital role in detection of diseases, however quality of images may be impacted due to noise. Therefore, to minimize the errors and human effort in medical imaging, an automated segmentation technique will provide better accuracy and ROI selection [23], [24], [25]. Deep learning models have been employed for various purposes to extract the efficient features i.e. medical [26], [27], agriculture [28], surveillance [29], etc. Although, the supervised methods provide better accuracy, however the challenging task is to label

the training samples of large number. Additionally, data may have different types, therefore to label and prepare the large training data is never-ending task.

The existing techniques are based on outdated machine learning algorithms which require hand-crafted feature extractions from images. Furthermore, they are unable to give high accuracy and precision rates for the detection of KOA. Whereas, deep learning-based algorithms for KOA require a huge training set size to perform significantly. They are unable to detect unseen images of Grade-I and Grade-II KOA disease precisely. Moreover, existing deep learning-based techniques performed better for binary classification such as healthy and diseased images of the Knee, however, the accuracy for individual grades of the disease according to the KL grading system is not considerable. Therefore, in the proposed study, we overcome the challenges of existing work using an improved deep learning model i.e. CenterNet with DenseNet201 as the base technique. We employed a customized CenterNet model with DenseNet-201 as a backbone network to compute the deep key points from input images, and localization of the knee joint as a region of interest based on the pixel-wise voting scheme. The proposed algorithm is based on two main phases such as training, and classification. Training of the proposed novel classifier has been performed over training samples using the Mendeley dataset. In the end, the classification phase has been performed on the images to classify them based on ROI. Furthermore, to enhance the performance of the proposed model, we employ knowledge distillation that can transfer valuable features and semantic knowledge from a complex teacher network to compact student network. The proposed algorithm is evaluated over a standard benchmark and results show that it achieves significant accuracy and outperformed the existing techniques.

The main contributions of the proposed study are:

- The main purpose of this work is to automate the KOA detection in an effective manner, therefore, a robust deep learning-based method i.e., CenterNet based on DenseNet-201 with pixel voting scheme is proposed that extracts the most representative features to detect and recognize early KOA effectively. We utilized weighted pixel-wise voting scheme for the best localization results from our customized CenterNet model.
- The proposed model is an efficient technique that provides precise localization and classification of the knee osteoarthritis according to the KL grading scheme due to its improved architecture of the CenterNet.
- We have introduced knowledge distillation in our modified network to reduce the size of our proposed model along with inference time without increasing extra computational power. Moreover, it also transfers the knowledge from the cumbersome model to a compact one.
- The model achieves 99.14% accuracy over testing and 98.97% over the OAI dataset during cross-validation.

It is concluded that our proposed model is robust enough to utilize over various knee samples.

- The proposed model is computationally fast and simple, based on a single-stage KOA detector and classifier.
- To assess the performance of the proposed model, extensive experiments have been performed which show that our algorithm outperformed the existing techniques.

II. LITERATURE REVIEW

The disease detection of the knee involves various methods for the examination of knee joints such as X-rays, MRI, and CT scans. KOA detection techniques can be categorized into three main types such as segmentation-based, feature extraction-based, and classification-based. One of the methods has been discussed in [15], however, it lacks the standard grading system for the assessment of KOA severity. An early KOA detection system has been introduced based on morphological features, mechanical and electrical properties, and molecular context [30]. Furthermore, the system is applicable for MRI images in non-ionized, in-vivo, and non-invasive modalities. In [31], authors used data from the Osteoarthritis Initiative (OAI) to analyze the progression of the disease. Steady State MRI with dual-echo has been used to assess the images, detect the region of comparison, and perform segmentation. Moreover, machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Networks (ANN) have been employed for comparison and selection of the best approach. Various segmentation-based techniques have been discussed in previous years [32], [33], [34]. Knee bone segmentation has been assessed in [32] and [35], whereas articular cartilage has been segmented only in [35], however, they didn't compare the proposed method quantitatively. Various deep learning-based architectures for bone segmentation and classification have been discussed in [36], [37], [38], [39], [40], and [41]. In [40], SegNet architecture is based on 10 layers without a fully connected layer after the decoder network has been developed for 2D knee images to employ semantic labeling pixel-wise. After this step, the segmented objects have been polished based on the original image. The model had fewer parameters due to the removal of the FC layer. Later, [41] modified the framework to compute numerous tissue segmentation using the conditional Random Forest (RF) for multi-classification. They have achieved 97% accuracy for the femur, 96.2% for the tibia, and 89.8% for the patella. [37] used the concept of segmentation based on the slice and added an extra feature of SSM in U-Net based segmentation framework. The model overcome the challenge of holes in segmentation masks as a result of poor intensity contrast and false-positive voxels that were identified outside the actual range. Although the model attained good accuracy, however, the computational cost was very high. To overcome the challenges, [38] developed a simple CNN-based technique i.e. Holistically Nested Network (HNN) for the ROI segmentation. HNN removed the

decoding path to perform a feed-forward network, therefore reducing the complexity of the model. In [36], authors have discussed various supervised machine learning (SML) methods' implications in the healthcare and biomedical sectors.

Deep learning techniques have shown remarkable results for the medical domain [42], [43]. In [44], a deep learning-based model has been proposed by authors to identify KOA disease considering the minimum-joint space width. The experiments showed that the proposed system significantly considers the knee joint space and classifies disease effectively. Wahid et al. [45], developed a multi-layered convolutional sparse model to categorize the MRI scan as an ACL tear but only for the coronal plane. Although it achieved a good accuracy of 85%, it was not of great use in diagnosis as it assessed only one of the three planes and only one type of injury. An effective object detection model is You Only Look at Once (YOLO) with CNNs can be used to localize the object (an area where the features reveal disease) [46]. Furthermore, a detailed review of existing techniques is reported in Table 1.

III. METHODOLOGY

In this work, a robust framework for the detection of KOA is suggested. The proposed system can be employed on unseen knee images having varying severity levels of KOA. The high-dimensional features play a significant role in the recognition and characterization of disease in knee images. We fed the samples having the annotated bounding boxes as a region of interest (ROI). We utilized improved CenterNet using DenseNet-201 as the base network for feature formation. The reason behind choosing the DenseNet over ResNet is to extract the most representative feature from the knee joint due to densely connected layers. However, ResNet employs skip connections and attains output from the second and third layers. Moreover, DenseNet contains a feature layer (convolutional layer) capturing low-level features from knee images, several dense blocks, and transition layers between adjacent dense blocks. Although, DenseNet requires high computational power, however, it provides better feature representation than ResNet.

Before, the feature extraction phase from the knee joint, we improved the localization results by giving input bounding box predicted from our customized CenterNet to the voting function. The voting function computes the best bounding box by taking votes from each pixel from the estimated bounding box and gives an output of the best bounding box based on maximum score. Additionally, to reduce the size of the model and to transfer the knowledge from a cumbersome model to a compact one without increasing the computational power, we have introduced knowledge distillation. Therefore, an automated model for the detection of KOA disease is employed using the dataset i.e. Mendeley. We trained an improved CenterNet network [55] over the various knee joints samples attained from the medical experts. Moreover, these samples are characterized rendering to the KL grading systems such as G-I, G-II, G-III, and G-IV. The architecture of the proposed system is shown in figure 1. After the training

TABLE 1. Details of some existing works.

Reference	Year	Dataset	Features	Algorithm	Validation	Results	Issues
Gan, H.-S., et al.[48]	2017	MRI	Fuzzy c-means and K-means Clustering	Seeds Labelling (Flexible)	10 images are validated by 2 experts	Dice's score of 80% by observer 1 and 82% by observer 2	Failed to give a significant performance on unseen data due to small training data used.
Kashyap, S., et al.[49]	2018	MRI	K-means clustering based on neighborhood approx. forests	LOGISMOS and Hierarchical RF Classifier	108 MRIs, and 12 months follow-up scans of 54 patients	Cartilage surface errors(mm) of 4D Femur signed 0.01 ± 0.18 Femur unsigned 0.53 ± 0.11 at baseline	It did not consider knee joints detailed features, and did not considered the severity of KOA.
Ambellan, F., et al.[37]	2019	MRI	-	Statistical Shape Models and 2D/3D CNN	SKI10, OAI Imorphics, OAI ZIB	$74\% \pm 7.7$ overall Score	Detection and classification results are not considerable.
Liu et al. [50]	2015	MRI	Automatic	Multi-atlas Context Forests	40 test images	$DSC \pm SD: 81.8 \pm 3.0\%$ (FC); $79.2 \pm 4.6\%$ (TC)	The method is based on only classification, therefore, considered the irrelevant features from knee images lead to degrade the performance.
Gornale, S.S., et al. [51]	2019	X-ray	Basic Statistical Features, ROI detection using Sobel, Prewitt and edge	KNN, Texture based segmentation, and Ostu's segmentation	532 Knee X-ray images	91.16% accuracy for Sobel, 96.80% for Ostu's, 94.92% for texture-based, and 97.55% for the Prewitt method is achieved.	The method took more computational time for training.
Gan et al. [52]	2019	MRI	Automatic	Random walks-SAGE	-	(OA): 93% (FC); 88%	The details of test data is not provided.
Cueva et al. [53]	2022	X-ray	Automatic	ResNet-34	376 images	Multiclass Accuracy: 61%	The detection accuracy for multi classification is very low.
[46]	2019	Demographic, personal features, lifestyle-related factors	PCA	DNN	34% for testing and 66% for training	76.8% AUC	The detection accuracy is very low.
[54]	2018	X-ray	-	SOM	8.8% and testing training 91.2%	40.52% accuracy for KL Grades-II and 36.21% for Grade-0	The detection and classification accuracy for Grade-II is very low.

of the classifier, classification is performed and images have been characterized into five classes i.e. Normal, G-I, G-II, G-III, and G-IV.

A. CENTERNET ARCHITECTURE

There exist various object detection techniques i.e. anchor-based detection techniques that create number of rectangles on the image to compute a maximum intersection over union (IoU) with labels. However, they encounter three issues such as: 1) the size of the anchor should be computed manually, 2) the number of rectangles requires high computational cost, and 3) the anchors don't need to always match with the ground truth bbox. Therefore, to cope with these challenges, Duan et al. [55] developed a mokey point CenterNet, based on

a key points object recognition pipeline. Moreover, CenterNet is based on CornerNet [56] that exhibits the object in form of paired corner key points and adds an extra key point at the center of the proposal. Therefore, each object is described by a pair of corners and center key point. The candidate bboxes are discarded by picking the top k center key points. These center key points are remapped to the input sample, then central regions are defined and checked if they consist of any center key point. Furthermore, CenterNet is more efficient as it specifies features and bboxes locations of objects in frames at the same time. Whereas, other object detection algorithms such as RCNN, Fast RCNN, and Faster RCNN perform detection in two stages that make them computationally inefficient for early KOA detection.

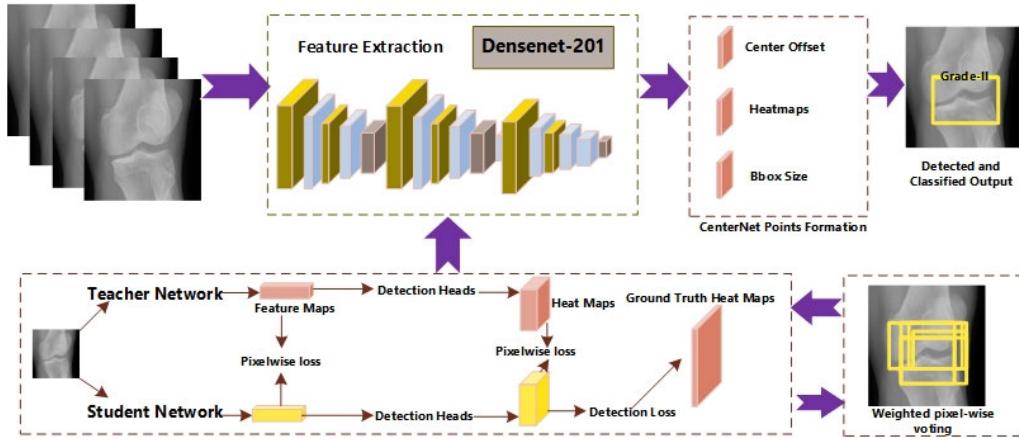


FIGURE 1. Architecture of the proposed model.

CenterNet detects the region of interest based on two main steps i.e. localization and size regression. It employs Gaussian Kernel for the localization and generates heat maps, therefore aiding the classifier to produce maximum activations near the center of objects. Whereas, for the regression, it expresses the pixels at the center of the object during training and then predicts the height and width of the object. It also forecasts the offset to overcome the discretization error due to the output stride R . Furthermore, after the computation of feature maps, three types of heat maps are generated: 1) Localization Heat maps, 2) Regression Heat maps, and 3) Offset Heat maps. Let $I \in R^{w \times h \times 3}$ is an input image having width w and height h and the bounding box of k is $(i_1^k, j_1^k, i_2^k, j_2^k)$ having category c . The key point $P^k \in \mathbb{R}^2$ refers to the center of an object can be described as in the below equation:

$$P^k = \left(\frac{i_1^k + i_2^k}{2}, \frac{j_1^k + j_2^k}{2} \right), \quad (1)$$

A key point's heat map is referred by $h_{ijc} \in \hat{w} \times \hat{h} \times c$, computed using 2D Gaussian kernel in equation 2.

$$\hat{h}_{ijc} = \exp \left(\frac{(i - p_i)^2 + (j - p_j)^2}{2\sigma_p^2} \right), \quad (2)$$

where $\hat{w} = \frac{w}{R}$, $\hat{h} = \frac{h}{R}$, $p = \lfloor \frac{p}{R} \rfloor$, R refers to the stride of output. C is the number of classes i.e. the expressions, and σ_p refers to the standard deviation of the object's size [56]. Algorithm 1 shows the steps of the proposed model.

B. BASE NETWORK

We employed DenseNet-201, is easy to train network and has less number of parameters. It provides the reusability of features using various layers increasing the variations in each input layer, therefore improving the accuracy [57]. Thus, the feature maps at each layer are concatenated with the previous layer's feature maps, therefore the channel's dimensions are increased at each layer. The concatenation of features is

represented mathematically in equation 3.

$$Y^l = h_l([y^0, y^1, \dots, y^{l-1}]), \quad (3)$$

where h_l presents a non-linear transformation composite function that consists of batch normalization (BN), rectified linear unit function (ReLU) [58], and convolution of 3×3 . Suppose, there exists l number of layers, the size of feature maps is k , and total channels are referred by M , then the output of the $l+1$ will be $k \times k \times (M + 2M)$. Dense Blocks are used for down-sampling and tangled with each other with the transition layers among them. Moreover, these transition layers are comprised of BN followed by a convolutional layer of 1×1 and a pooling layer of 2×2 . 1×1 convolutional layer is preceded by a 3×3 convolutional layer to improve the computational efficiency. These layers are called the bottleneck layers, which decrease the input feature maps if they are more than the output feature maps. The extracted feature maps are down-sampled at the final transition layer having a stride of 2. DenseNet consists of four dense blocks as shown in figure 2. The details of layers of DenseNet-201 and Resnet-101 are shown in Table 2.

1) FOCAL LOSS

The localization loss of CenterNet's key point is a modified form of focal loss [59], significantly overcoming the problem of class imbalance between positive and negative class frames. Thus, it is unable to deal with the different class's imbalance that reduces the accuracy of object detection in unusual categories. Therefore, we employ a category-based focal loss function that overcomes both issues such as category and positive and negative class imbalance in the training samples. The loss can be defined mathematically as below.

$$l_s = -\frac{1}{n} \sum_{ijc} \times w_c \begin{cases} (1 - h_{ijc})^\alpha \log(h_{ijc}), & \text{if } h_{ijc} = 1 \\ (1 - h_{ijc})^\beta (h_{ijc})^\alpha \log(1 - (1 - h_{ijc})), & \text{Otherwise,} \end{cases} \quad (4)$$

TABLE 2. Details of Densenet-201 and Resnet-101 layers.

Proposed Backbone Network: Densenet-201			Original Backbone Netowrk of CenterNet: Resnet-101		
Layers	Output Size	No. of Units	Layers	Output Size	No. of Units
Convolution	112 x 112	7 x 7 Conv. and Stride =2	Convolution 1	112 x 112	[7 * 7 Conv, Stride =2 Channel 64] x 1
Pooling	56 x 56	3 x 3 max pool and Stride=2			[3 * 3 max pooling, Stride=2] x 1
Dense Block 1	56 x 56	[1 * 1 conv] x 6 [3 * 3 conv]	Convolution 2	56 x 56	[1 * 1, Channel 64] x 3 [3 * 3, Channel 64] x 3
Transition Layer 1	56 x 56	1 x 1 conv			[1 * 1, Channel 256] x 3
	28 x 28	2 x 2 avg. pooling and Strdie=2	Convolution 3	28 x 28	[1 * 1, Channel 128] x 4 [3 * 3, Channel 128] x 4
Dense Block 2	28 x 28	[1 * 1 conv] x 12 [3 * 3 conv]			[1 * 1, Channel 512] x 4
Transition Layer 2	28 x 28	1 x 1 conv	Convolution 4	14 x 14	[1 * 1, Channel 256] x 23 [3 * 3, Channel 256] x 23
	14 x 14	2 x 2 avg. pooling and Strdie=2			[1 * 1, Channel 1024] x 23
Dense Block 3	14 x 14	[1 * 1 conv] x 48 [3 * 3 conv]	Convolution 5	7 x 7	[1 * 1, Channel 512] x 3 [3 * 3, Channel 512] x 3
Transition Layer 3	14 x 14	1 x 1 conv			[1 * 1, Channel 2048] x 3
	7 x 7	2 x 2 avg. pooling` Strdie=2	Output	1 x 1	Average Pooling, 1000-d fc, Softmax
Dense Block 4	7 x 7	[1 * 1 conv] x 32 [3 * 3 conv]		-	-
Classification Layer	1 x 1	7 x 7 global avg. pool FC softmax		-	-

Here α and β represent the hyper-parameters for the function, n refers to the total key points in Image I , and h_{ijc} represents the localization of key points output. To compute the weights of class method has been employed as:

$$w_c = \frac{W - 1}{s_c^\sigma} (s_c^\sigma - 1) + 1, \quad (5)$$

Here $s_c = \frac{K_{max}}{K_c}$, where K_c is the number of boxes that are labeled for class c , K_{max} , and K_c represent the maximum and minimum numbers respectively. W and σ present the hyper-parameters, whereas 1 refers to the most frequent and W to the rarest class.

The discretization error due to the output stride is minimized by the offset prediction branch to set the position of the center point and all the classes have the same offset prediction phase. The offset loss is shown mathematically below.

$$l_{off} = \frac{1}{N} \sum \left| \hat{O}_p - \left(\frac{P}{R} - p \right) \right|, \quad (6)$$

Here $\hat{O}_p \in \hat{W} \times \hat{h} \times 2$ represents the offset prediction of each central point. To revert the size of object $S_k = (i_2^k - i_1^k, j_2^k - j_1^k)$, we employed prediction branch for size to attain $\hat{R} \in \hat{R} \hat{W} \times \hat{h}^2$

for all categories. This branch was trained with the loss at center as:

$$l_{size} = \frac{1}{n} \sum_{i=1}^n |M_{pk} - M_k|, \quad (7)$$

The training objective function for the detection of objects is computed as below:

$$L_{tot} = l_s + \lambda_{off} l_{off} + l_{size} \lambda_{size}, \quad (8)$$

Here λ_{off} and λ_{size} represent the tunable parameters for the balancing of all loss functions.

C. KNOWLEDGE DISTILLATION

To enhance the performance of the proposed model, we employ knowledge distillation that can transfer valuable features and semantic knowledge from complex teacher network to compact student network i.e., ResNet26 comprising of 85 layers. This module made our model simple and increased the robustness of the proposed method as student network was trained over mixed labels such as original labels and outputs from teacher network. We employed the concept of teacher and student network similar to [60] to transfer

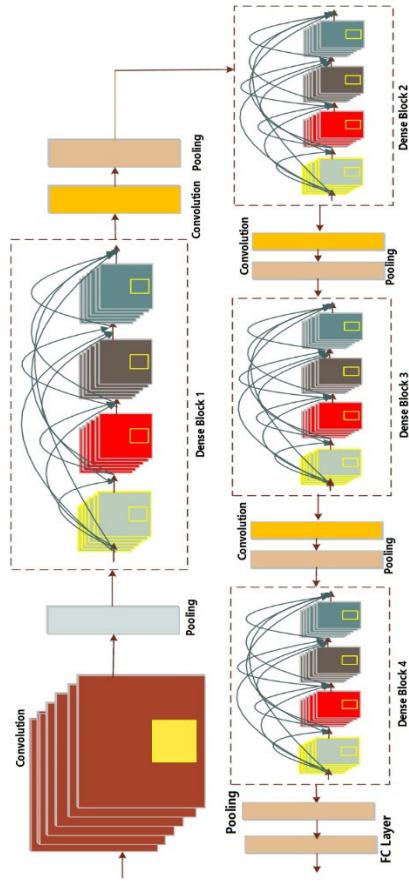


FIGURE 2. Architecture of DenseNet-201.

the pixel and pair-wise knowledge from teacher to student network. Knowledge distillation

The base of teacher and student networks are HRNet-W(32,16) respectively. The teacher network architecture has head convolution layer having 64 filters along with 3×3 kernel, whereas student network has 32 filters having 3×3 kernels. The pixel-wise knowledge utilizes heat maps attained from the cumbersome network to train the simple and compact network. The loss function is denoted as below:

$$L_{pi} = \frac{\sum_{i \in \mathfrak{N}} KL(h_i^s || h_i^t)}{\hat{w} \times \hat{h}}, \quad \mathfrak{N} = 1, 2, \dots, \hat{w} \times \hat{h}, \quad (9)$$

where h_i^s is the response of the pixel at i th position in the student network s , whereas h_i^t refers to the response by the teacher network t at i th position, and KL represents the kullback leibler exhibiting divergence among two heat maps. Moreover, the similarity index is computed between two pixels that is employed using cosine. Suppose, feature map has $\hat{w} \times \hat{h} \times c$, here c refers to the total channels and $\hat{w} \times \hat{h}$ is the feature map size, and $f_i \in \mathbb{R}^c$ is a feature vector attained from the spatial position i of feature map. Therefore, the cosine similarity \mathbb{Q}_{ij} is computed as below:

$$\mathbb{Q}_{ij} = \frac{f_i^t f_j}{\|f_i\|_2 \cdot \|f_j\|_2}, \quad (10)$$

Algorithm 1 Steps for Proposed Model

```

Input: Training Images (TI), annotations
Output: Localized ROI consisting of Knee joints, Classified Disease
Start:
Image_size  $\leftarrow [i,j]$ 
 $\mathbb{C}' \leftarrow \text{EstimatedAnchors}(TI, \text{annotations})$  // Bounding Box Estimation
 $\mathbb{C} \leftarrow \text{Weighted\_pixel\_wise\_voting}$ 
CD  $\leftarrow \text{DenseNet based CenterNet (Image\_size, } \mathbb{C} \text{)}$  // CenterNet
 $[I_{train}, I_{test}] \leftarrow \text{Separation of training and testing knee samples}$ 
// Training for early knee disease detection
For  $\forall$  image  $x$  in  $\rightarrow I_{train}$ 
    Computation of keypoints  $\rightarrow T_s$ 
End For
Train CD for  $T_s$ , calculation of training time t_time
n_pos  $\leftarrow \text{KneeJointLocation}(T_s)$ 
E_dense  $\leftarrow \text{Compute\_Accuracy (DenseNet, n_pos)}$ 
WHILE  $\forall I \in I_{test}$ 
    1.  $\dot{\eta} \leftarrow \text{Extraction of frames through trained}$ 
    network @
        2.  $bbox, ClassLabel, Detection\_Score \leftarrow \dot{\eta}$ 
        3. Show sample having  $bbox$ , and  $ClassLabel$ 
    End While
    Accuracy computation to assess the model's performance
End
```

The pair wise distillation loss is given below. \mathbb{Q}_{ij}^t is the similarity index of the teacher network and \mathbb{Q}_{ij}^s is the similarity index of the student network.

$$L_{pa} = \frac{\sum_{i \in \mathfrak{N}} \sum_{j \in \mathfrak{N}} (\mathbb{Q}_{ij}^s - \mathbb{Q}_{ij}^t)^2}{(\hat{w} \times \hat{h})^2}, \quad \{\mathfrak{N} = 1, 2, \dots, \hat{w} \times \hat{h}\}, \quad (11)$$

Therefore, the below is the resulting loss function:

$$L_{KD} = L_{tot} + \lambda_{pa} L_{pa} + \lambda_{pi} L_{pi}, \quad (12)$$

where λ_{pa} and λ_{pi} are tunable parameters for the loss functions.

Thus, an improved CenterNet is a deep learning-based independent model that doesn't require proposal generation. Thus, the input sample with bbox is fed to the trained classifier to compute the center key points of knee joints as ROI, the x and y coordinate offsets along with the dimensions and related class. The localization heatmaps are shown in Figure 3.

D. WEIGHTED PIXEL-WISE VOTING

For the improvement of the bounding box localization, we employed a pixels-based voting scheme. The voting method generates the summarized predictions for the bounding box considering all the pixels belonging to an input box b in form of improved bounding box predictions such as

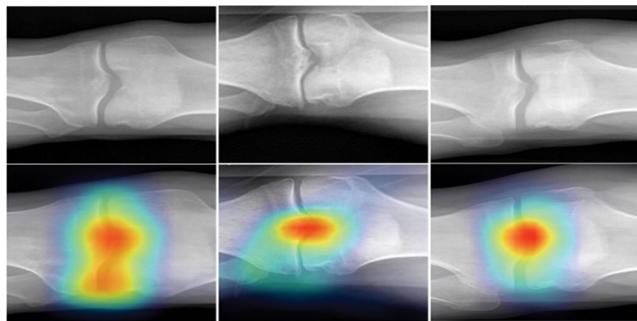


FIGURE 3. Samples localization heatmaps.

(i^b, j^b, h^b, w^b) such as:

$$i^b, j^b, h^b, w^b = V_f(b), \quad (13)$$

Here, V_f presents the voting function. To enhance the localization results, we used a histogram of votes employing parameters of the bounding box, in which the estimation having maximum value for confidence $c_{k,l}$ takes part using larger weights. More particularly, we calculate the score of voting for the target knee joint box from the center as $B_{center}^b \in R^{h*w}$, height as $B_{height}^b \in R^h$, and width as $B_{width}^b \in R^w$ from the input box b . The mathematical notations are shown below:

$$B_{center}^b(i, j) = \sum_{k, l \in b} c_{k, l} \cdot II([k + r_{k, l} \sin(\theta_{k, l})] = j) \cdot II([l + r_{k, l} \cos(\theta_{k, l})] = i), \quad (14)$$

$$B_{width}^b(w) = \sum_{k, l \in b} c_{k, l} \cdot II([w_{k, l}] = w), \quad (15)$$

$$B_{height}^b(h) = \sum_{k, l \in b} c_{k, l} \cdot II([h_{k, l}] = h), \quad (16)$$

Here, $II(\cdot)$ is exhibiting the indicating function and $\lfloor \dots \rfloor$ is referring to rounding a number to the nearest integer. Moreover, $r_{k, l}$ presents the reward value in the context of reinforcement learning and $\theta_{k, l} \in [0, 2\pi]$ refers the estimated displacement relatively from (k, l) to the respective bounding box center in accordance with polar coordinate systems. $h_{k, l} \in (0, H]$ is the height and $w_{k, l} \in (0, W]$ is the width of the bounding box estimation at (k, l) . Furthermore, $c_{k, l} \in [0, 1]$ exhibits the confidence value estimation at (k, l) . The optimized parameters for bounding boxes are (i^b, j^b, h^b, w^b) and achieved by searching the candidate having maximum V_f in B_{center}^b , B_{width}^b , and B_{height}^b respectively.

In order to recognize the knee joint accurately, the pixel-wise voting scheme is utilized. The function is developed in a way that individual pixel in the predicted input bounding box by improved CenterNet gives a vote for an improved BBox. In the end, box having the maximum votes is selected as the output box. Although, one round for voting is not worthy for accurate localization of knee joint. Therefore, we employed various iterations using supervised learning till we get significant results of knee joint localization.

IV. EXPERIMENTAL EVALUATION

This section describes the extensive details of the experiments and results for the analysis of the proposed model. More

TABLE 3. Details of hyper-parameters.

Hyper-parameters	Value
Batch size	64
Learning rate	0.0001
Step size	50k, 80k
Decay rate	10
NMS threshold	0.5
Optimized algorithm	1500

precisely, in section IV-A implemented experiment setup is described and in IV-B dataset detail is presented, and. In sections IV-C - IV-G, metrics and various experiments have been reported that we employed to assess the performance of the proposed technique.

A. EVALUATION SETUP AND METRICS

The experiment was performed on a Windows-based system with Intel® Xeon(R) at 3.30 GHz × 4, 32 GB RAM having GPU card as NVIDIA GEFORCE GTX × 4. Moreover, the proposed model was implemented using the Python framework and library of Keras version: v0.1.1. The hyper-parameters were set as 500 epochs, learning rate: 0.0001, and batch size: 64. The details of hyper-parameter is given in Table 3 .

B. DATASETS

We used two datasets for various experiments i.e. for training, testing, and cross-validation. The dataset namely Mendeley Data V1 [61] is widely used for KOA severity detection and classification according to the KL grading system. It is comprised of 2000 digital knee X-ray images of 8-bits and 1350×2455 dimensions. For the evaluation, the knee images have been annotated by two medical experts to separate them according to the KL grading scheme. Moreover, images were grayscale and in PNG format. We divided the dataset as 75% for training and validation, and 25% for testing. More precisely, the model has trained over 1600 images and tested over 400 images of the knee. We considered the combined grading by both Experts for the distribution of datasets such as 500 images from the Healthy class, 400 images from Grade-I, 200 images from Grade-II, 200 images from Grade-III, and 200 images from Grade-IV class have been used for the training. The class-wise distribution of knee images for training and testing of the proposed model is presented in figure 4. The details of train and test samples are presented in TABLE 4. Some samples from the dataset have been shown in figure 5.

Furthermore, we have employed the Osteoarthritis Initiative (OAI) dataset for cross validation. It consists of 3T MRI scans and X-rays for knee joints having the categorization according to KL grading systems. The data was collected from 4,796 participants including males and females of age 45-79. Moreover, patients who had knee replacement surgery

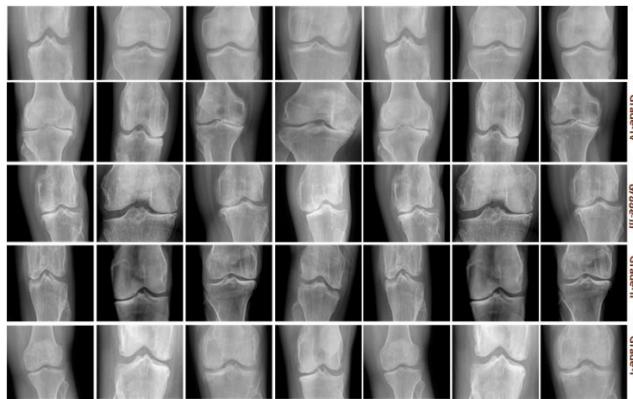


FIGURE 4. Samples from dataset.

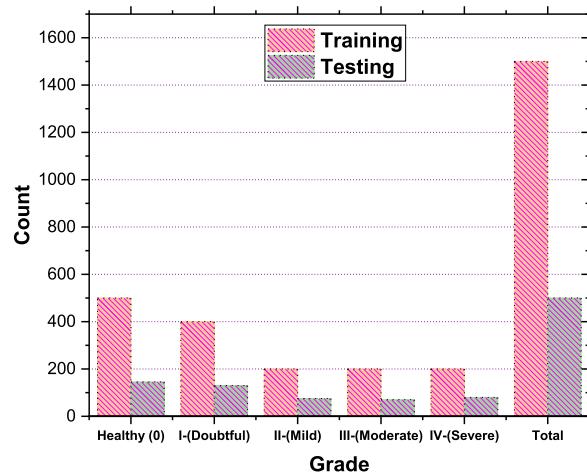


FIGURE 5. Plot for class-wise distribution of dataset.

TABLE 4. Summary of mendeley dataset.

Grade	Training	Testing
Healthy (0)	500	145
I-(Doubtful)	400	130
II-(Mild)	200	75
III-(Moderate)	200	70
IV-(Severe)	200	80
Total	1500	500

were not included in the dataset. Among 4,796 participants, 5,045 cases were of the healthy knee and 3,967 images belonged to grades II, III, and IV. Additionally, we tested our proposed model over 2500 images which include 500 images of each class i.e. Healthy, Grade-I, Grade-II, Grade-III, and Grade-IV.

C. METRICS

For the assessment of the suggested model, we have used various metrics i.e., Precision, Recall, Accuracy, F1 Score, and dice similarity coefficient (DSC). Moreover, these metrics are relied on true positive (TP), true negative (TN), false positive (FP), and false-negative (FN). The TP denotes correctly classified images by our proposed model, FP refers

to the images that were incorrectly categorized as a positive class, FN denotes the diseased images that were incorrectly categorized as normal, and TN refers the number of images that were correctly classified as a negative class. Furthermore, precision refers to the fraction of TP over the total images classified as positive. The mathematical equation is shown below.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (17)$$

Accuracy of the system indicates the correctly classified images by the proposed system. The equation is presented below.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (18)$$

The recall is the fraction of the classified positive class images to all images of positive class whether they were classified as a negative class by the system. The recall value closer to 1 refers to the better model. The equation of Recall is given below.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (19)$$

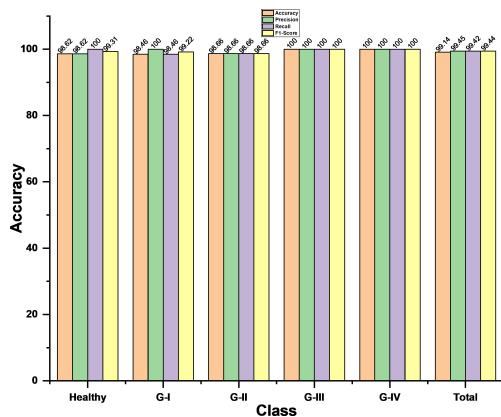
Moreover, another metric used for the proposed system was the F1 Score. It is defined as a measure of the accuracy of the proposed model over the dataset. It is employed for binary classification models. The equation of the F1 Score is given below.

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (20)$$

In the end, we employed dice similarity coefficient (DSC) for evaluation that is commonly used for medical imaging segmentation. It is compared with the ground truth values attained by the medical expert. A refers to the medical expert and V is the volume. The mathematical formulae for the DSC is described below.

$$\text{DSC} = \frac{2V_A \cap V_B}{V_A + V_B}, \quad (21)$$

The results of our proposed model are reported in Table 7. It can be seen that total of 500 images have been tested for evaluation. Moreover, for each class, a detailed confusion matrix is presented such as among 145 total healthy images, 143 images have been correctly categorized as a healthy class given as TP, 2 images have been incorrectly classified as FP i.e. diseased class, and accuracy attained for this class is 98.62%. Similarly, among 130 G-I images, 128 have been classified as G-I and only 2 images have been classified as healthy class (FN) attaining 98.46% accuracy. For G-II among 75 images only one image has been incorrectly classified as FP i.e. G-I class. Furthermore, for G-III, and G-IV all images have been correctly classified and we attained 100% accuracy. Therefore, collectively our proposed algorithm attains 99.14% accuracy, 99.45% precision, 99.42% recall, 99.44% F1-score, and 99.24% DSC with 0.03 standard deviation. Our proposed algorithms perform better than our previous work [19] for KOA detection. Here, we achieve significant performance for all grade detection

**FIGURE 6.** Performance plot over Mendeley dataset.**TABLE 5.** Ablation study: performance with losses.

Loss(mIOU%)	Baseline Model	DenseNet 201	Inception V2	MobileNet V1
Heatmap loss	12.1	2.3	13.2	9.3
Offset loss	9.4	1.7	8.5	9.9
Object loss	14.2	2.3	11.2	8.4

due to implication of segmentation and localization that considers pixel features more precisely. Further, the accuracy for Grade-I is 98.47% exhibiting that our proposed system identifies the early KOA effectively, consequently preventing the disease progression to severe stages. The performance plot over Mendeley dataset is shown in figure 6.

D. ABLATION STUDY

The contributions of our work are a concatenation of a pixel-wise voting scheme with CenterNet to improve the performance of detection and localization along with knowledge distillation. Moreover, we employed DenseNet as the base network of CenterNet, which has direct connections among layers, therefore extracting the most representative features. We compared the results of our proposed model with the baseline model, InceptionV2 and MobileNetV1 as base networks. In TABLE 5, it can be seen that after employing our proposed pixel-wise voting scheme, losses have decreased significantly. It clearly conveys that by using DenseNet201, our technique has significantly reduced the losses of CenterNet while improving the localization and classification result.

We performed another experiment to analyze the performance of the knowledge distillation function. The experiment is conducted on ResNet variants such as ResNet8, ResNet20, ResNet26 and ResNet18 on the Mendeley dataset. We used Teacher accuracy, student accuracy, and layers as param-

TABLE 6. Knowledge distillation results for different teacher models.

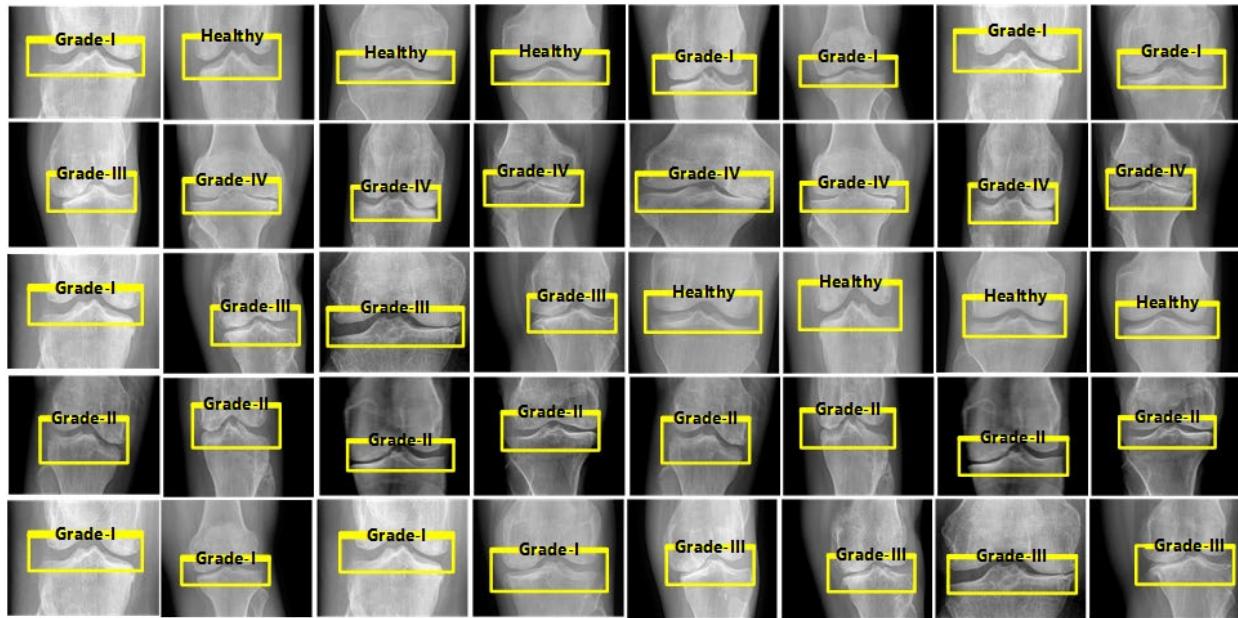
Teacher	ResNet8	ResNet20	ResNet26	ResNet18
Layers	31	67	85	62
Teacher Accuracy	91.2	93.04	94.3	89.2
Student Accuracy	91.4	90.2	97.4	88.4
Model's accuracy	92.3	94.5	99.14	91.7

ters. As shown in Table 6, we attained the best results for the Teacher's model using ResNet26. Moreover, the overall accuracy of the proposed system has been increased to 99.14% from 94.7% without knowledge distillation. Thus, introducing the knowledge distillation, the network's robustness has also increased.

E. COMPARISON WITH THE EXISTING DL MODELS

In this section, we compare our proposed model with the existing deep learning models for the detection and classification of KOA disease. Most of the Techniques have employed the OAI dataset for the experimental evaluation. In [46], the Deep CNN model has been used for KOA detection using 62,419 images taken from the Institutes in South Korea. They have achieved 76.8% accuracy with a training time of 266.67 seconds and a testing time of 14.63 seconds. In [16], authors have implemented CNN to assess the knee images and used two different datasets i.e. OAI, and MOST. The training time of the algorithm was 423.67 seconds and the testing time was 15.74 seconds. They have achieved 63.40% accuracy for the classification. Authors in [10] employed OAI, and MOST datasets using Siamese Deep NN and achieved 66.71% accuracy, which is not a considerable one. The algorithm utilized 150.67 seconds for training, and 3.8 seconds for testing. Although, the training and testing times are less than the aforementioned techniques, however, the results are not accurate. Some test images attained from the proposed model implication are shown in figure 7.

Moreover, in [62] and [63], the OAI dataset has been used for the experiments. They have achieved an accuracy of 69.70%, 75.28%, and 77.24% respectively. The training and testing time was as 364.67, and 25.43 seconds for [63], 510, and 39.37 seconds for [64], and 457, and 28.53 seconds for the [62]. Furthermore, our proposed algorithm has employed the Mendeley dataset for training and testing, and the OAI dataset for cross-validation. It attained 99.14% accuracy for the Mendeley test set and 98.7% accuracy for the OAI dataset. In addition, the training time was 156 seconds and the inference time was 2.3 seconds that is the minimum time for the experiments than existing DL models. Our proposed

**FIGURE 7.** Test samples of Mendely dataset.**TABLE 7.** Performance analysis of the proposed model upon testing.

Class	Total Images	TP	TN	FP	FN	Accuracy (%)	Precision (%)	Recall (%)	F1-Score(%)	DSC (%)
Healthy	145	143	0	2	0	98.62	98.62	100	99.31	99.4±0.03
G-I	130	128	0	0	2	98.46	100	98.46	99.22	98.9±0.12
G-II	75	74	0	1	0	98.66	98.66	98.66	98.66	98.7±0.02
G-III	70	70	0	0	0	100	100	100	100	99.6±0.001
G-IV	80	80	0	0	0	100	100	100	100	99.6±0.001
Total	500	491	0	4	5	99.14	99.45	99.42	99.44	99.24±0.03

TABLE 8. Comparison with existing DL models.

Reference	Year	Dataset	Algorithm	Accuracy (%)	Training Time(s)	Testing Time(s)
[46]	2019	62,419 images	DCNN	76.8	266.67	14.63
[16]	2017	OAI(4,446), MOST(2920)	CNN	63.40	423.67	15.74
[10]	2018	OAI(3000), MOST	Siamese DNN	66.71	150.67	3.8
[62]	2020	OAI	DCNN	77.24	457	28.53
[63]	2019	OAI	DCNN	69.70	364.67	25.43
[64]	2020	OAI	LSTM	75.28	510	39.37
[65]	2022	OAI	ComplexNet	81.69	-	-
Our Model	2022	Mendeley, OAI	DenseNet as base of Centernet	99.14	156	2.3

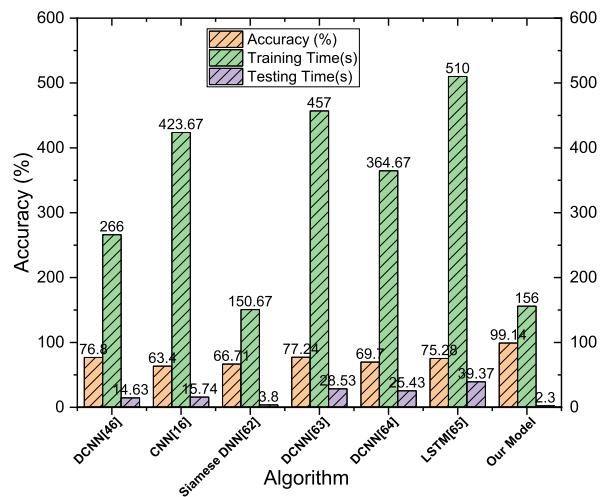
algorithm is robust and categorize images efficiently due to its dense architecture. It can be seen in table 8, that our proposed algorithm outperformed all existing techniques in terms of accuracy, robustness, and training and testing time. Comparative plot is shown in figure 8.

F. CROSS-VALIDATION

Here, we perform experiment to assess the robustness of our proposed model using the Osteoarthritis Initiative (OAI) dataset. The data was collected from 4,796 participants including males and females of age 45-79. Moreover, patients

TABLE 9. Cross-validation results over OAI dataset.

Class	Total Images	TP	TN	FP	FN	Accuracy (%)	Precision (%)	Recall (%)
Healthy	700	69	0	8	0	98.86	98.85	100
Grade-I	600	59	0	0	9	98.5	100	98.5
Grade-II	600	59	0	7	0	98.83	98.83	100
Grade-III	600	59	0	0	2	99.66	100	99.6
Grade-IV	500	49	0	5	0	99	99	100
Total	3000	29	0	2	1	98.97	99.36	99.63

**FIGURE 8.** Comparison with existing DL models.

who had knee replacement surgery were not included in the dataset. Among 4,796 participants, 5,045 cases were of the healthy knee and 3,967 images belonged to grades II, III, and IV. Furthermore, we tested our proposed model over 3000 images which include images of each class i.e. 700 for Healthy, 600 for G-I, 600 for G-II, 600 for G-III, and 500 for G-IV. The performance results are reported in Table 9. It is depicted that the proposed model attains 98.97% accuracy, 99.36% precision, and 99.63% recall. More explicitly, among 700 healthy knee images, 8 knee images have been incorrectly classified, and among 600 G-I images only 9 images have been incorrectly classified as healthy knee images due to very minor differences between Healthy and G-I images. Similarly among 600 G-II, G-III, and 500 G-IV knee images, 593, 598, and 495 images have been correctly classified respectively. Henceforth, the proposed algorithm effectively localizes the knee joint and classifies the knee images into five classes i.e. Healthy, G-I, G-II, G-III, and G-IV. The Grade-wise accuracies are 98.86%, 98.5%, 98.83%, 99.66%, and 99% respectively. It can be seen that our proposed algorithm is giving the best results employing DenseNet-201 as a base network.

G. DISCUSSION

Our proposed system is designed for orthopedic surgeons or radiologists, who can use it for early detection of KOA and its severity using X-rays. The system aids in the prevention of the severe stage of KOA disease by detecting it in the early stage i.e., Grade-I. Moreover, from the experiments, it is clear that our KOA detector is robust enough which can identify disease in unseen images effectively. Typically the isolation of the tibiofemoral joint (ROI) is a tedious task due to its irregular boundaries and manually performing the segmentation is quite difficult. Therefore, the proposed model identifies the ROI based on the voting scheme effectively and then extracts the most representative features due to the dense connections of DenseNet as a base network. Feature extraction from the region of interest based on dense blocks has been incorporated to improve the predictive capabilities of the proposed system and also with the aim to increase its detection accuracy. Hence, health organizations can easily adopt our proposed system for the detection of KOA severity using X-rays. For example, if a patient feels mild pain due to any reason in his knee and gets his X-rays from a radiologist. Then, the proposed system can be used by the radiologists to inform the patient that he doesn't have developed KOA. Eventually, it will save time and effort of patients and Orthopedics as well.

V. CONCLUSION

In this study, we propose a robust deep learning architecture to detect Knee Osteoarthritis (KOA) and identify severity levels based on KL grading i.e. G-I, G-II, G-III, and G-IV. The proposed system is based on an improved CenterNet architecture using Densenet-201 that effectively overcomes the challenge of class imbalance in the dataset. Densenet-201 extracts the most representative features from the identified ROI due to dense connections among all layers. Moreover, we employed the distillation knowledge to make our model simple without increasing its computational cost and transfer knowledge from a complex network to a simple network making it more robust. Besides this, to improve the localization results for knee joint detection, we used a weighted pixel-wise voting scheme that detects the knee joint region accurately. We utilized two datasets in the proposed study such as: 1) Mendeley Dataset used for training and testing, and 2) OAI Dataset used for cross-validation. Various experiments have been performed to assess the performance of the proposed model. The model attained 99.14% accuracy for the overall localization of knee joints and classification of KOA. In addition, 98.97% accuracy has been achieved over cross-validation on the OAI dataset, which exhibits the robustness of the proposed algorithm. The most important key aspect of our study is to utilize it for the early and accurate localization of knee joints and detection of the KOA according to the KL grading system and minimize the time and cost needed for the additional examination approaches. Thus, the proposed technique can easily be employed in healthcare organizations for early diagnosis of KOA. Raw X-ray images can be fed

into our proposed procedure directly and all adjustments are handled internally. Moreover, it can also be used for other disease detection, and categorization such as tumor detection. We noticed that our proposed method consumed huge time for training. Therefore, in the future, we aim to reduce the training time while simplifying the network. Additionally, we will apply this method to other fields such as plant disease detection and emotion analysis.

ACKNOWLEDGMENT

The authors would like to thank the Deanship of Scientific Research, Qassim University for funding publication of this Project.

REFERENCES

- [1] T. Tsonga, M. Michalopoulou, P. Malliou, G. Godolias, S. Kapetanakis, G. Gkasdaris, and P. Soucacos, "Analyzing the history of falls in patients with severe knee osteoarthritis," *Clinics Orthopedic Surg.*, vol. 7, no. 4, pp. 449–456, 2015.
- [2] B. J. E. de Lange-Brokaar, A. Ioan-Facsinay, E. Yusuf, A. W. Visser, H. M. Kroon, S. N. Andersen, L. Herb-van Toorn, G. J. V. M. van Osch, A.-M. Zuurmond, V. Stojanovic-Susulic, J. L. Bloem, R. G. H. H. Nelissen, T. W. J. Huizinga, and M. Kloppenburg, "Degree of synovitis on MRI by comprehensive whole knee semi-quantitative scoring method correlates with histologic and macroscopic features of synovial tissue inflammation in knee osteoarthritis," *Osteoarthritis Cartilage*, vol. 22, no. 10, pp. 1606–1613, Oct. 2014.
- [3] C. Kokkotis, C. Ntakolia, S. Moustakidis, G. Giakas, and D. Tsaopoulos, "Explainable machine learning for knee osteoarthritis diagnosis based on a novel fuzzy feature selection methodology," *Phys. Eng. Sci. Med.*, vol. 45, no. 1, pp. 219–229, Mar. 2022.
- [4] S. Nalband, R. R. Sreekrishna, and A. A. Prince, "Analysis of knee joint vibration signals using ensemble empirical mode decomposition," *Proc. Comput. Sci.*, vol. 89, pp. 820–827, Jan. 2016.
- [5] B. J. Guo, Z. L. Yang, and L. J. Zhang, "Gadolinium deposition in brain: Current scientific evidence and future perspectives," *Frontiers Mol. Neurosci.*, vol. 11, p. 335, Sep. 2018.
- [6] 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 Trans. Biomed. Eng.*, vol. 56, no. 2, pp. 407–415, Feb. 2009.
- [7] A. Brahim, R. Jennane, R. Riad, T. Janvier, L. Khedher, H. Toumi, and E. Lespessailles, "A decision support tool for early detection of knee OsteoArthritis using X-ray imaging and machine learning: Data from the OsteoArthritis initiative," *Comput. Med. Imag. Graph.*, vol. 73, pp. 11–18, Apr. 2019.
- [8] P. S. Emrani, J. N. Katz, C. L. Kessler, W. M. Reichmann, E. A. Wright, T. E. McAlindon, and E. Losina, "Joint space narrowing and Kellgren-Lawrence progression in knee osteoarthritis: An analytic literature synthesis," *Osteoarthritis Cartilage*, vol. 16, no. 8, pp. 873–882, Aug. 2008.
- [9] M. N. Iqbal, F. R. Haidri, B. Motiani, and A. Mannan, "Frequency of factors associated with knee osteoarthritis," *J. Pakistan Med. Assoc.*, vol. 61, no. 8, p. 786, 2011.
- [10] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, and S. Saarakkala, "Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, Jan. 2018.
- [11] M. S. M. Swamy and M. S. Holi, "Knee joint cartilage visualization and quantification in normal and osteoarthritis," in *Proc. Int. Conf. Syst. Med. Biol.*, Dec. 2010, pp. 138–142.
- [12] P. Dodin, J. Pelletier, J. Martel-Pelletier, and F. Abram, "Automatic human knee cartilage segmentation from 3-D magnetic resonance images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 11, pp. 2699–2711, Nov. 2010.
- [13] N. Kour, S. Gupta, and S. Arora, "A survey of knee osteoarthritis assessment based on gait," *Arch. Comput. Methods Eng.*, vol. 28, no. 2, pp. 345–385, Mar. 2021.
- [14] M. Saleem, M. S. Farid, S. Saleem, and M. H. Khan, "X-ray image analysis for automated knee osteoarthritis detection," *Signal, Image Video Process.*, vol. 14, no. 6, pp. 1079–1087, Sep. 2020.
- [15] J. Abedin, J. Antony, K. McGuinness, K. Moran, N. E. O'Connor, D. Rebholz-Schuhmann, and J. Newell, "Predicting knee osteoarthritis severity: Comparative modeling based on patient's data and plain X-ray images," *Sci. Rep.*, vol. 9, no. 1, pp. 1–11, Apr. 2019.
- [16] J. Antony, K. McGuinness, K. Moran, and N. E. O'Connor, "Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks," in *Proc. 13th Int. Conf. Mach. Learn. Data Mining Pattern Recognit. (MLDM)*. New York, NY, USA: Springer, Jul. 2017, pp. 376–390.
- [17] J. Antony, K. McGuinness, N. E. O'Connor, and K. Moran, "Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 1195–1200.
- [18] F. R. Mansour, "Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy," *Biomed. Eng. Lett.*, vol. 8, no. 1, pp. 41–57, Feb. 2018.
- [19] R. Mahum, S. U. Rehman, T. Meraj, H. T. Rauf, A. Irtaza, A. M. El-Sherbeeny, and M. A. El-Meligy, "A novel hybrid approach based on deep CNN features to detect knee osteoarthritis," *Sensors*, vol. 21, no. 18, p. 6189, Sep. 2021.
- [20] C. Cernazanu-Glavan and S. Holban, "Segmentation of bone structure in X-ray images using convolutional neural network," *Adv. Electr. Comput. Eng.*, vol. 13, no. 1, pp. 87–94, 2013, doi: [10.4316/AECE.2013.01015](https://doi.org/10.4316/AECE.2013.01015).
- [21] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. B. Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," *Comput. Methods Programs Biomed.*, vol. 104, no. 3, pp. e158–e177, Dec. 2011.
- [22] C. Stoljescu-Crișan and S. Holban, "A comparison of X-ray image segmentation techniques," *Adv. Electr. Comput. Eng.*, vol. 13, no. 3, pp. 85–92, 2013.
- [23] H. S. Gan and K. A. Sayuti, "Comparison of improved semi-automated segmentation technique with manual segmentation: Data from the osteoarthritis initiative," *Amer. J. Appl. Sci.*, vol. 13, no. 11, pp. 1068–1075, 2016, doi: [10.3844/ajassp.2016.1068.1075](https://doi.org/10.3844/ajassp.2016.1068.1075).
- [24] Y. Li, N. Xu, and Q. Lyu, "Construction of a knee osteoarthritis diagnostic system based on X-ray image processing," *Cluster Comput.*, vol. 22, no. S6, pp. 15533–15540, Nov. 2019.
- [25] S. Kubkaddi and K. Ravikumar, "Early detection of knee osteoarthritis using SVM classifier," *Int. J. Sci. Eng. Adv. Technol.*, vol. 5, no. 3, pp. 259–262, 2017.
- [26] Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, and Y. Liu, "Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI," *Inf. Fusion*, vol. 91, pp. 376–387, Mar. 2023.
- [27] A. Adegun and S. Viriri, "Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art," *Artif. Intell. Rev.*, vol. 54, no. 2, pp. 811–841, Jun. 2020.
- [28] M. A. Guillén, A. Llanes, B. Imbernón, R. Martínez-España, A. Bueno-Crespo, J.-C. Cano, and J. M. Cecilia, "Performance evaluation of edge-computing platforms for the prediction of low temperatures in agriculture using deep learning," *J. Supercomput.*, vol. 77, no. 1, pp. 818–840, Jan. 2021.
- [29] B. Janakiramaiah, G. Kalyani, and A. Jayalakshmi, "Retraction note: Automatic alert generation in a surveillance systems for smart city environment using deep learning algorithm," *Evol. Intell.*, vol. 14, no. 2, pp. 635–642, Dec. 2022.
- [30] A. F. M. Hani, A. S. Malik, D. Kumar, R. Kamil, R. Razak, and A. Kiflie, "Features and modalities for assessing early knee osteoarthritis," in *Proc. Int. Conf. Electr. Eng. Informat.*, Jul. 2011, pp. 1–6.
- [31] A. E. Nelson, F. Fang, L. Arbeeva, R. J. Cleveland, T. A. Schwartz, L. F. Callahan, J. S. Marron, and R. F. Loeser, "A machine learning approach to knee osteoarthritis phenotyping: Data from the FNIH biomarkers consortium," *Osteoarthritis Cartilage*, vol. 27, no. 7, pp. 994–1001, Jul. 2019.
- [32] A. Aprovitola and L. Gallo, "Knee bone segmentation from MRI: A classification and literature review," *Biocybern. Biomed. Eng.*, vol. 36, no. 2, pp. 437–449, 2016.
- [33] V. Pedoia, S. Majumdar, and T. M. Link, "Segmentation of joint and musculoskeletal tissue in the study of arthritis," *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 29, no. 2, pp. 207–221, Apr. 2016.
- [34] J. Kubicek, M. Penhaker, M. Augustynek, I. Bryjova, and M. Cerny, "Segmentation of knee cartilage: A comprehensive review," *J. Med. Imaging Health Informat.*, vol. 8, no. 3, pp. 401–418, Mar. 2018.
- [35] B. Zhang, Y. Zhang, H. D. Cheng, M. Xian, S. Gai, O. Cheng, and K. Huang, "Computer-aided knee joint magnetic resonance image segmentation—A survey," 2018, *arXiv:1802.04894*.

- [36] T. Meena and S. Roy, "Bone fracture detection using deep supervised learning from radiological images: A paradigm shift," *Diagnostics*, vol. 12, no. 10, p. 2420, Oct. 2022.
- [37] S. Roy, T. Meena, and S.-J. Lim, "Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine," *Diagnostics*, vol. 12, no. 10, p. 2549, Oct. 2022.
- [38] D. Pal, P. B. Reddy, and S. Roy, "Attention UW-Net: A fully connected model for automatic segmentation and annotation of chest X-ray," *Comput. Biol. Med.*, vol. 150, Nov. 2022, Art. no. 106083.
- [39] H. Lee, H. Hong, and J. Kim, "BCD-NET: A novel method for cartilage segmentation of knee MRI via deep segmentation networks with bone-cartilage-complex modeling," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1538–1541.
- [40] F. Liu, Z. Zhou, H. Jang, A. Samsonov, G. Zhao, and R. Kijowski, "Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging," *Magn. Reson. Med.*, vol. 79, no. 4, pp. 2379–2391, 2018.
- [41] Z. Zhou, G. Zhao, R. Kijowski, and F. Liu, "Deep convolutional neural network for segmentation of knee joint anatomy," *Magn. Reson. Med.*, vol. 80, no. 6, pp. 2759–2770, Dec. 2018.
- [42] R. Mahum, H. Munir, Z.-U.-N. Mughal, M. Awais, F. S. Khan, M. Saqlain, S. Mahamad, and I. Tlili, "A novel framework for potato leaf disease detection using an efficient deep learning model," *Hum. Ecol. Risk Assessment, Int. J.*, vol. 29, no. 2, pp. 303–326, Feb. 2023.
- [43] S. Sikandar, R. Mahmum, and N. Akbar, "Cricket videos summary generation using a novel convolutional neural network," in *Proc. Mohammad Ali Jinnah Univ. Int. Conf. Comput. (MAJICC)*, Oct. 2022, pp. 1–7.
- [44] J. C.-W. Cheung, A. Y.-C. Tam, L.-C. Chan, P.-K. Chan, and C. Wen, "Superiority of multiple-joint space width over minimum-joint space width approach in the machine learning for radiographic severity and knee osteoarthritis progression," *Biology*, vol. 10, no. 11, p. 1107, Oct. 2021.
- [45] A. Wahid, J. A. Shah, A. U. Khan, M. Ullah, and M. Z. Ayob, "Multi-layered basis pursuit algorithms for classification of MR images of knee ACL tear," *IEEE Access*, vol. 8, pp. 205424–205435, 2020, doi: [10.1109/ACCESS.2020.3037745](https://doi.org/10.1109/ACCESS.2020.3037745).
- [46] Y. Wang, X. Wang, T. Gao, L. Du, and W. Liu, "An automatic knee osteoarthritis diagnosis method based on deep learning: Data from the osteoarthritis initiative," *J. Healthcare Eng.*, vol. 2021, pp. 1–10, Sep. 2021.
- [47] M. S. Swanson, J. W. Prescott, T. M. Best, K. Powell, R. D. Jackson, F. Haq, and M. N. Gurcan, "Semi-automated segmentation to assess the lateral meniscus in normal and osteoarthritic knees," *Osteoarthritis Cartilage*, vol. 18, no. 3, pp. 344–353, Mar. 2010.
- [48] H.-S. Gan, K. A. Sayuti, N. H. Harun, and A. H. A. Karim, "Flexible non cartilage seeds for osteoarthritic magnetic resoano image of knee: Data from the osteoarthritis initiative," in *Proc. IEEE EMBS Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2016, pp. 748–751.
- [49] 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 Trans. Med. Imag.*, vol. 37, no. 5, pp. 1103–1113, May 2018.
- [50] Q. Liu, Q. Wang, L. Zhang, Y. Gao, and D. Shen, "Multi-atlas context forests for knee MR image segmentation," in *Proc. 6th Int. Workshop Mach. Learn. Med. Imag. (MLMI)*, Munich, Germany: Springer, Oct. 2015, pp. 186–193.
- [51] S. S. Gornale, P. U. Patravali, A. M. Uppin, and P. S. Hiremath, "Study of segmentation techniques for assessment of osteoarthritis in knee X-ray images," *Int. J. Image, Graph. Signal Process.*, vol. 11, no. 2, pp. 48–57, Feb. 2019.
- [52] H.-S. Gan, K. A. Sayuti, M. H. Ramlee, Y.-S. Lee, W. M. H. W. Mahmud, and A. H. A. Karim, "Unifying the seeds auto-generation (SAGE) with knee cartilage segmentation framework: Data from the osteoarthritis initiative," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 5, pp. 755–762, May 2019.
- [53] J. H. Cueva, D. Castillo, H. Espinós-Morató, D. Durán, P. Díaz, and V. Lakshminarayanan, "Detection and classification of knee osteoarthritis," *Diagnostics*, vol. 12, no. 10, p. 2362, Sep. 2022.
- [54] L. Anifah, M. H. Purnomo, T. L. R. Mengko, and I. K. E. Purnama, "Osteoarthritis severity determination using self organizing map based Gabor kernel," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 306, Feb. 2018, Art. no. 012071.
- [55] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6569–6578.
- [56] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 734–750.
- [57] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
- [58] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*.
- [59] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [60] Y. Liu, K. Chen, C. Liu, Z. Qin, Z. Luo, and J. Wang, "Structured knowledge distillation for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2604–2613.
- [61] S. S. Gornale, P. U. Patravali, and P. S. Hiremath, "Automatic detection and classification of knee osteoarthritis using Hu's invariant moments," *Frontiers Robot. AI*, vol. 7, Nov. 2020, Art. no. 591827.
- [62] R. T. Wahyuningrum, A. Yasid, and G. J. Verkerke, "Deep neural networks for automatic classification of knee osteoarthritis severity based on X-ray images," in *Proc. 8th Int. Conf. Inf. Technol., IoT Smart City*, Dec. 2020, pp. 110–114.
- [63] P. Chen, L. Gao, X. Shi, K. Allen, and L. Yang, "Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss," *Comput. Med. Imag. Graph.*, vol. 75, pp. 84–92, Jul. 2019.
- [64] R. T. Wahyuningrum, L. Anifah, I. K. E. Purnama, and M. H. Purnomo, "A new approach to classify knee osteoarthritis severity from radiographic images based on CNN-LSTM method," in *Proc. IEEE 10th Int. Conf. Awareness Sci. Technol. (ICAST)*, Oct. 2019, pp. 1–6.
- [65] L. C. Ribas, R. Riad, R. Jennane, and O. M. Bruno, "A complex network based approach for knee osteoarthritis detection: Data from the osteoarthritis initiative," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103133.

SULIMAN ALADHADH received the B.Sc. degree in computer science from Qassim University, Saudi Arabia, in 2008, and the M.Sc. degree in advance certificate in computer science research and the Ph.D. degrees in computer science from RMIT University, Australia, in 2012, 2014, and 2019, respectively. He is currently an Assistant Professor with the College of Computer, Qassim University. His research interests include data science and social computing.

RABBIA MAHUM received the B.Sc. degree in computer science from COMSATS University Islamabad, Wah Campus, in 2015, and the M.S. degree in computer science from UET Taxila, in 2018, where she is currently pursuing the Ph.D. degree. Her research interests include computer vision, deep fake audio detection, and medical imaging.