Computers in Biology and Medicine 167 (2023) 107570

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

DC-AAE: Dual channel adversarial autoencoder with multitask learning for KL-grade classification in knee radiographs

Muhammad Umar Farooq [a], Zahid Ullah [b], Asifullah Khan [c], Jeonghwan Gwak [a,b,d,e,*]

[a] Department of IT, Energy Convergence (BK21 FOUR), Korea National University of Transportation, Chungju 27469, South Korea
[b] Department of Software, Korea National University of Transportation, Chungju 27469, South Korea
[c] Pattern Recognition Lab, DCIS, PIEAS, Nilore, Islamabad 45650, Pakistan
[d] Department of Biomedical Engineering, Korea National University of Transportation, Chungju 27469, South Korea
[e] Department of AI Robotics Engineering, Korea National University of Transportation, Chungju 27469, South Korea

ARTICLE INFO

Keywords:
Knee osteoarthritis
Multi-task learning
Deep learning
Semi-supervised learning
KL-grade classification

ABSTRACT

Knee osteoarthritis (OA) is a frequent musculoskeletal disorder that leads to physical disability in older adults. Manual OA assessment is performed via visual inspection, which is highly subjective as it suffers from moderate to high inter-observer variability. Many deep learning-based techniques have been proposed to address this issue. However, owing to the limited amount of labelled data, all existing solutions have limitations in terms of performance or the number of classes. This paper proposes a novel fully automatic Kellgren and Lawrence (KL) grade classification scheme in knee radiographs. We developed a semi-supervised multi-task learning-based approach that enables the exploitation of additional unlabelled data in an unsupervised as well as supervised manner. Specifically, we propose a dual-channel adversarial autoencoder, which is first trained in an unsupervised manner for reconstruction tasks only. To exploit the additional data in a supervised way, we propose a multi-task learning framework by introducing an auxiliary task. In particular, we use leg side identification as an auxiliary task, which allows the use of more datasets, e. g., CHECK dataset. The work demonstrates that the utilization of additional data can improve the primary task of KL-grade classification for which only limited labelled data is available. This semi-supervised learning essentially helps to improve the feature learning ability of our framework, which leads to improved performance for KL-grade classification. We rigorously evaluated our proposed model on the two largest publicly available datasets for various aspects, i.e., overall performance, the effect of additional unlabelled samples and auxiliary tasks, robustness analysis, and ablation study. The proposed model achieved the accuracy, precision, recall, and F1 score of 75.53%, 74.1%, 78.51%, and 75.34%, respectively. Furthermore, the experimental results show that the suggested model not only achieves state-of-the-art performance on two publicly available datasets but also exhibits remarkable robustness.

1. Introduction

Osteoarthritis (OA) is the most common form of the joint disease affecting up to 18.0% of women and 9.6% of men over the age of 60 around the world [1]. OA is generally characterized by pain, swelling, stiffness, and a grating sensation on movement [2–4]. OA is a degenerative joint disease caused by an irreparable deterioration of the joint cartilage located at the end of the bone [5]. These symptoms severely affect the quality of elderly people's life. Knee OA is one of the leading health burdens contributing to chronic disability in the world [6,7]. Further, there is a no-disease specific drug to date for this irreversible degenerative disease [8]. Medication only assists in reducing OA symptoms, specifically pain. It has been reported that by the year 2050, this degenerative joint disease disorder (OA) will affect about 130 million people around the world [9]. Considering the rapid increase in the number of knee OA patients, it is paramount to detect knee OA at the initial stage and started proper treatment.

Most of the existing techniques for knee OA diagnosis are based on plain radiography because of their easy accessibility and availability [10]. Moreover, through radiography images, the most interesting diagnostic features such as subchondral bone changes, joint space narrowing (JSN), and osteophyte formation could be visualized vividly. However, such practice allows the diagnosis of knee OA at a later stage when the cartilage is already worn, and the bone deformity is present, which takes an individual to physical disability [2,3]. Therefore, in this scenario, total knee replacement (TKR) surgery is the only option for a patient. Previous studies illustrate a worldwide growing number of

* Corresponding author at: Department of Software, Korea National University of Transportation, Chungju 27469, South Korea.
E-mail address: james.han.gwak@gmail.com (J. Gwak).

https://doi.org/10.1016/j.compbiomed.2023.107570
Received 4 May 2022; Received in revised form 25 August 2023; Accepted 10 October 2023
Available online 13 October 2023

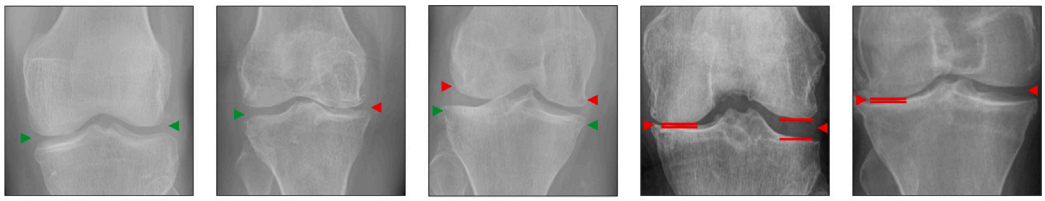

Fig. 1. KL-grade classification for knee OA severity grading. (a) Normal: No radiological features of Osteoarthritis, (b) Doubtful: Possible narrowing of joint space and/or doubtful osteophytes, (c) Mild: Definite osteophytes and possible narrowing of joint space, (d) Moderate: Definite narrowing of joint space, and some sclerosis and deformity of bone ends, (e) Severe: Large osteophytes, severe sclerosis, and definite deformity of bone ends.

TKR surgeries [11,12]. In the United States, the annual rate of TKR surgeries in adults aged 45–64 years has doubled since 2000 [13,14]. Hence, there is a need for the prevention of global disability.

Imaging, as compared to clinical examination, may enable the early detection of knee OA signs at the periods when behavioural interventions (for instance, weight loss and exercise programs) could slow down the progress of the disease [15]. The Radiographic assessment is the major imaging tool for knee OA detection in primary care, while Kellgren and Lawrence (KL) are commonly used to evaluate the severity of knee OA from plain radiographs. The KL presented the first radiographic classification criteria, based on these features, to grade and identify OA [16]. There are total of five KL-grade scores such as 0 for the normal case to 4 for a severe case, which is being used to score knee OA as shown in Fig. 1. However, visual diagnosis performed by a radiologist suffers from low inter-rater agreement [17], which generates ambiguity and makes the diagnosis of knee OA very challenging. Leveraging a computer-aided diagnosis system is the only possible solution to make OA diagnosis more systematic and also allow knee OA detection at early stages [18]. Recently, Deep Learning (DL)-based techniques have garnered attention in various industrial applications due to their impressive performance [19–21]. These applications span object classification [22–25], segmentation [26–29], medical image motion correction [30,31], and object detection [32]. Specifically, DL-based techniques have outperformed traditional machine learning techniques for automatic KL-grading and have become the state-of-art approach in this realm [17,33,34].

In the OA domain, most of the previously published DL methods employ supervised learning techniques, which require a sufficiently large labelled dataset. However, at present, the labelled data is not extensively available. Alternatively, unlabelled data can be employed, which can be easily acquired from hospital imaging archives at a significantly lower cost. Therefore, several recent studies [35–37] have employed semi-supervised learning (SSL) for solving labelled dataset problems by employing both labelled data and much smaller unlabelled data. However, such studies utilized unlabelled data only in an unsupervised manner, which restrains the model from learning critical information, and thus leads to substandard performance.

Alternatively, multi-task learning (MTL) [38] is a more effective approach to enhancing the generalization of the model, which can simultaneously learn relevant features for the auxiliary task along with the primary task. MTL also helps in the better regularization of models, which eases the exploration of the common high-level discriminative features. MTL has already been applied for various problems, such as speech emotion recognition [39], COVID-19 detection [40], and tumour detection [41], and has attained significant performance improvement. Similarly, we exploit the MTL for the KL-grade classification task and introduce a dummy auxiliary task for which data annotation is easy, and thus helps in enabling the utilization of all the unlabelled data during the supervised learning process. Specifically, the identification of leg sides, i.e., the left and right sides, is introduced as the auxiliary task for MTL.

Furthermore, we use generative adversarial models within our MTL framework because of their exceptional ability to learn the most efficient features [42]. Specifically, we utilize an adversarial autoencoder (AAE) [43], which mainly learns the representation of data in an unsupervised way. After combining AAE with the supervised classification networks, we then enable SSL for AAE. Mainly, the unsupervised training pipeline incorporates the discriminative component of supervised learning to support the latent representation of AAE, thus preparing the grounds for semi-supervised KL-grade classification.

The proposed MTL framework is comprehensively evaluated via two widely used and large datasets (i.e., osteoarthritis initiative (OAI) and multicenter osteoarthritis study (MOST) datasets). A detailed comparison has been provided based on parameters like accuracy, precision, recall, and F1-score to evaluate the effectiveness of the proposed approach in terms of overall performance and robustness. Moreover, a detailed analysis of the effects of incorporating additional data and the auxiliary task has been performed.

The main contributions of our proposed work are summarized as follows:

- Aiming to develop a novel semi-supervised multi-task learning-based technique for automatic knee KL-grade classification that helps to exploit additional unlabelled data in an unsupervised as well as supervised manner.
- Introducing a novel dual-channel adversarial autoencoder (DC-AAE) for automatic KL-grading of knee OA from plain radiographs. The dual-channel helps the network to focus on the knee joints, which enables our method to outperform the existing single-encoder autoencoder (SE-AE).
- DC-AAE has been evaluated on OAI and MOST datasets, extensive experiments with different aspects such as overall performance, robustness analysis, effects of additional, and multitask learning effects. The results show DC-AAE not only outperforms the state-of-the-art method but also has generalization power which makes our model robust during cross-data analysis.

The organization of this research is laid out as follows. Section 2 presents related work and the background. A detailed description of the proposed technique has been given in Section 3. While Section 4 is dedicated to the experimental setup, The findings, and discussion are presented in Section 5. The ablation study is discussed in Section 6. Finally, Section 7 concludes this paper.

2. Related work

Knee OA severity assessment through classification is possible to achieve by detecting the osteophytes formation in the knee joints and variations in joint space width [44–46]. To consider the OA detection and KL-grading and to understand it deeply, numerous review papers have been carried out where they mentioned significant contributions [47–49]. Moreover, Yoo et al. [50] presented a scoring system to predict radiographic and symptomatic knee OA risks using artificial neural networks and KNHANES V-1 data.

Recently, CNN outperformed many methods that are based on handcrafted features [51]. In numerous computer vision tasks, CNN-based techniques are highly successful, such as video classification [52],

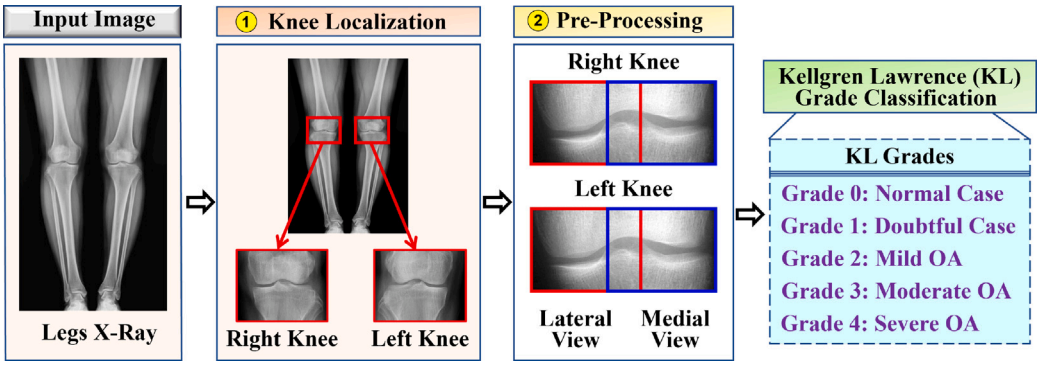

Fig. 2. Overall pipeline of the proposed KL-grade classification solution. In stage one, the bilateral legs X-ray image is fed and the left and right legs are localized. In step two, the knee joint images are centrally cropped with twenty percent overlapping. Lastly, our proposed DC-AAE model performs KL-grade classification for both knees.

image recognition [53], content-based image retrieval [54], and automatic detection and segmentation tasks [55,56]. CNN learns efficient feature representations that are more suitable for fine-grained classification [57] like knee OA image classification. Furthermore, the visual assessment process of radiographs is tedious. Therefore, different techniques for automatic diagnosis of knee OA severity, based on DL have recently been developed [17,58–60], which demonstrates that DL-based methods allow fully automatic KL-grading. Nguyen et al. [58] proposed a semixup algorithm to leverage unlabelled data and assess the KOA severity. They have considered two publicly available datasets, OAI and the MOST, for model training and testing, respectively, and achieved reasonable results.

Aleksei et al. [17] proposed a CNN-based model called Deep Siamese CNN, which minimizes the number of learnable parameters as compared to standard CNN. Their reported method automatically detects the severity of KOA according to the KL-grading scale. They used random seeds to select different knee joint sub-regions and then fused the predictions of the selected regions and obtained 66.71% accuracy. Shamir et al. [46] presented a template matching method with a sliding window strategy to locate the knee joints. They computed the Euclidean distances between 20 pre-defined knee joint images and the current shifted window on the down-scaled knee X-ray image. The smallest Euclidean distance window is considered as the detected knee joint. Chen et al. [59] developed a novel ordinal loss function to train deep neural networks to measure knee OA severity. They utilized a customized one-stage YOLOv2 network to detect the knee joint, and then the knee KL assessment was performed using the VGG-19 model and achieved 69.7% accuracy.

Albert et al. [61] proposed a fully-automated DL algorithm to assess the severity of knee OA in radiographs using the KL-grading system. The proposed model consists of two major steps, such as the localization of the knee joints in the images and KL-grading classification. They evaluated their work using the MOST dataset and obtained 71.90% accuracy. Antony et al. [34] presented a fully convolutional network (FCN) based approach to detect knee joint regions and classify the severity of different stages of knee OA and achieved good performance. Nasser et al. [62] introduced an architecture called discriminative regularized autoencoder to detect and classify the early detection of knee OA such as grade 0, grade 1, and grade 2. They have combined a discriminative loss with the standard AE training criterion that aims to force the learned representation to consider discriminative information. This model is evaluated on the OAI dataset and obtained state-of-the-art performance.

Anthony et al. [63] determine the OA severity level by utilizing transferred learning and the classical VGG-16 CNN architecture with X-ray images. They preprocessed these images using the Sobel edge detector and SVM to locate the knee joint area. The model is evaluated on the OAI dataset, which consists of 4,446 X-rays, representing 8,892 total knees. They achieved an accuracy of 59.6% when classifying the five KL-grades, which is low as compared to the state-of-the-art accuracy. Therefore, the same group [34] tried to improve the accuracy by updating the preprocessing step to utilize a FCN to determine the knee joint's bounding box. The FCN method, when combined with a CNN for classification, achieved an accuracy of 61.9%, which illustrates that the FCN method is highly effective in determining the region of interest (ROI). They evaluated their proposed method using OAI and MOST datasets, which comprised 4,446 images and 5,840 knee images, respectively, and hence improved the precision rate and accuracy rate of detecting knee joints.

3. Proposed method

In each bilateral X-ray scan, we localized each knee joint region and leg side selection using a publicly available tool [64] which is employed to detect the ROI i.e., images containing left and right knee joints. The overall flow diagram of the proposed architecture is depicted in Fig. 2, whereas Section 2 depicts the algorithm pipeline. After cropping the ROI around the knee joint, we further cropped the knee joint image by excluding the 25% region from the top and bottom sides, as shown in Fig. 3. We extracted 60% of the residual image from both horizontal sides, i.e., left and right sides, to obtain medial and lateral views. To avoid the loss of critical spatial contextual information during the central cropping of the knee joint, we performed the lateral and medial cropping with 20% overlapping from the middle of the knee joint. Finally, the lateral and medial views of the knee joint are obtained to feed into the proposed DC-AAE architecture.

Our proposed technique mainly consists of KL-grade prediction which is performed using DC-AAE networks. The details of each step are provided in the following subsections.

3.1. Preprocessing

The data used in this study, i.e., OAI dataset, was in DICOM file format which provides the window width (WW) and window centre (WC) dicom tags. The values of these tags are used to determine the range of CT (Computed Tomography) values or pixel intensities that will be displayed by default. The values of these tags are typically set by the radiologist or the technologist during image acquisition or post-processing, depending on the clinical task and the anatomical region being imaged. It is important to note that the window width can affect the diagnostic accuracy of an image and should be chosen carefully to optimize the visualization of the ROI. We perform central cropping on the knee joint to obtain medial and lateral views. The output of the tool [64] is cropped horizontally by 50% from the middle, followed by central cropping with 20% overlapping from the middle of the knee joint vertically. We cropped the knee joint image by excluding 25% regions from the top and bottom sides, as shown in Fig. 3. To obtain medial and lateral views, we extracted 60% of the residual image from

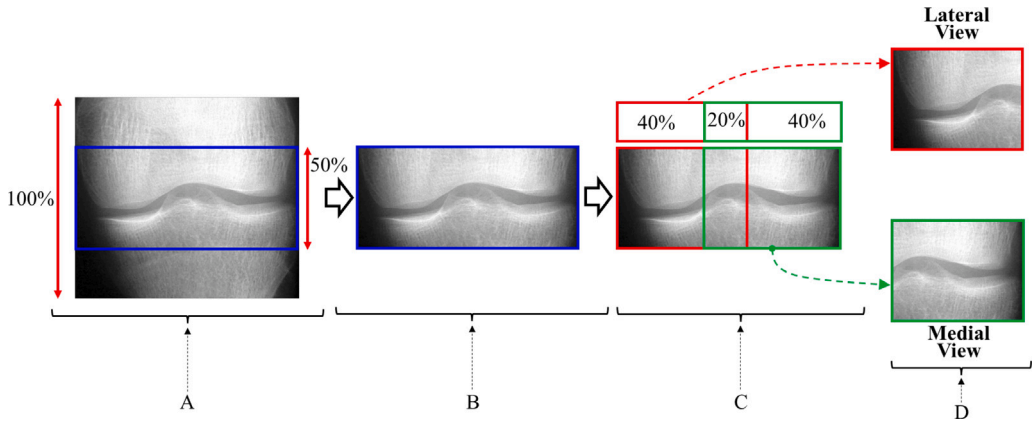

Fig. 3. Pipeline of the central cropping process, (A) is an input image, (B) is 50% image cropped horizontally from the middle, (C) is the central cropping with 20% overlapping from the middle of the knee joint vertically, and (D) shows the Lateral and Medial views of the knee joint.

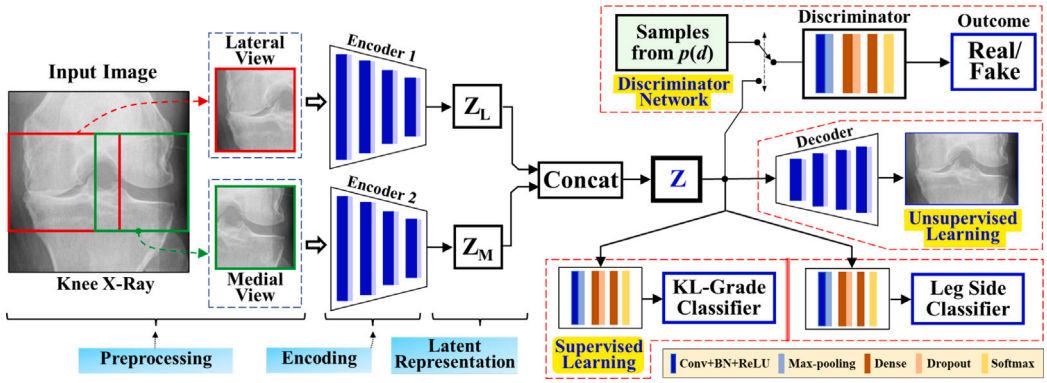

Fig. 4. Structure of DC-AAE with multitask learning for KL-grade classification in the knee radiographs.

both horizontal sides, i.e., the left and right sides. To avoid the loss of critical spatial contextual information due to central cropping of the knee joint, we performed the lateral and medial cropping with 20% overlapping from the middle of the knee joint. Finally, the lateral and medial views of the knee joint are obtained to feed into the proposed DC-AAE architecture. Fig. 4 depicts the proposed model flow diagram.

We also normalized the intensity values, ranging from 0 to 1, by using the window centre and window width tag from corresponding DICOM files [65]. The normalization can be defined as follows:

$$I_n = \frac{I - W_{Min}}{W_{Max} - W_{Min}}, \tag{1}$$

$$W_{Min} = WC - WW/2, \tag{2}$$

$$W_{Max} = WC + WW/2, \tag{3}$$

where I, I_n, WC, and WW represent the original image, normalized image, window centre, and window width, respectively. The values of the window centre and window width are extracted from the DICOM tags [65].

3.2. KL-grade classification using DC-AAE

We proposed a multi-task learning-based dual-channel adversarial autoencoding using an adversarial autoencoder (DC-AAE) framework which also incorporates semi-supervised learning. A typical AAE consists of a traditional autoencoder and an adversarial network to enhance the latent representation learning ability of the framework. In AAE, the encoder acts as a generator and generates latent code z which is connected to the adversarial part (i.e., discriminator). Here, the discriminator enforces the generator to produce latent representation z with similar statistical properties of a given prior distribution $p(d)$. To further enhance the learning ability of typical AAE, we introduced two encoders that generate two latent codes (i.e., z_L and z_M) from the lateral and medial parts of the knee. The discriminator enforces distribution by concatenating both latent representations to create z. Notably, the role of the discriminator in an AAE is to distinguish real (prior distribution) and fake (generated by the encoder). The discriminator is trained on prior distribution because it provides a well-defined and consistent target for the discriminator to learn.

In order to exploit the MTL in DC-AAE for utilization of additional data in supervised learning, we modified it by incorporating two supervised classification networks, including KL-grade and leg-side classification. Our proposed semi-supervised multitask learning model has been illustrated in Fig. 4, where the supervised and unsupervised paths are highlighted. Eq. (4) describes the multi-task autoencoding loss $\mathcal{L}_{\text{MTAE}}$ which is the combination of supervised and unsupervised losses.

$$\mathcal{L}_{\text{MTAE}} = \alpha * \mathcal{L}_{\text{AE}} + \mathcal{L}_C, \tag{4}$$

$$\mathcal{L}_C = \beta * \mathcal{L}_{\text{KL}} + (1 - \beta) * \mathcal{L}_{\text{LS}}. \tag{5}$$

Here, $\mathcal{L}_{\text{AE}}$ is the reconstruction loss of the autoencoder as shown in Eq. (6); in Eq. (5) $\mathcal{L}_{\text{KL}}$ and $\mathcal{L}_{\text{LS}}$ are losses for the KL-grade and leg side classification tasks, respectively; to control the weight of each loss term, we use α and β as the trade-off parameters.

DC-AAE contains the autoencoding network (encoders (E_θ) and decoder (D_δ)) and classifiers (C_ϕ), there is an adversarial network that includes a generator (E_θ) and discriminator (D_ω). We trained the model in three phases by using input data x: the first reconstruction (unsupervised) phase, the second regularization phase, and finally the

classification phase. The encoders (E_θ) and the decoder (D_δ) are updated during the reconstruction phase to reduce the reconstruction error by encoding x into latent representation z. The autoencoder's objective function is defined as follows:

$$\mathcal{L}_{\mathrm{AE}}(x, D_\delta(E_\theta(x))) = \|x - \hat{x}\|_2^2 \tag{6}$$

The proposed MTL framework is comprehensively evaluated via two widely used and large datasets (i.e., OAI and MOST datasets). A detailed comparison has been provided based on parameters like accuracy, precision, recall, F1-score, and AUC values to evaluate the effectiveness of the proposed approach in terms of overall performance and robustness. Moreover, a detailed analysis of the effects of incorporating additional data and the auxiliary task has been performed. We used an adversarial discriminator as a regularization method to improve the quality of the learned representation. During the regularization phase, the adversarial network first improves its ability to differentiate between samples from the actual distribution [real] and those generated by the encoder's latent representation. Then the encoder improves the ability to produce samples that the discriminator believes as real. This is achieved by maintaining the discriminator's weight and bias and adjusting the encoder's weight and bias through error backpropagation. The equation used to define the discriminator's objective is shown in Eq. (7).

$$\begin{aligned} \mathrm{L}_{\mathrm{disc}} = \max_{\omega} \big(& \mathrm{E}_{d\sim p_d}[\log(D_\omega(d))] \\ & + \mathrm{E}_{x\sim p_x}[\log(1 - D_\omega(E_\theta(x)))] \big) . \end{aligned} \tag{7}$$

Here p_x is the data distribution and $p(d) = \mathcal{N}(d; 0, I)$ is the prior multivariate Gaussian distribution.

In the *classification phase*, the latent code ($z = E_\theta(x)$) is input to classifiers (C_ϕ), which minimizes the standard class entropy loss by using the target vector containing the labels of both tasks and predicted values. Backpropagation of error is performed to update E_θ. The following objective function is used to optimize the encoder/generator network (E_θ).

$$\begin{aligned} \mathrm{L}_{\mathrm{enc}} = \min_{\theta} \big(& \mathrm{E}_{x\sim p_x}[\log(1 - D_\omega(E_\theta(x)))] \\ & + \mathrm{E}_{x\sim p_x}[\beta \mathcal{L}_{\mathrm{AE}}(x, D_\delta(E_\theta(x)))] + \\ & \mathrm{E}_{x,y\sim p_{X,Y}}[\mathcal{L}_c(E_\theta(x), y; C_\phi)] \big) . \end{aligned} \tag{8}$$

The encoder is updated in all three phases, while the discriminator is trained only once. The first, second, and third terms in Eq. (8) are updated in the regularization, reconstruction, and classification phases, respectively. Furthermore, all these phases are executed sequentially: first, the reconstruction phase runs, and after that, the regularization and classification phases are executed. Subsequently, the latent representation (z generated by the encoder during the unsupervised phase, is further fine-tuned by the supervised classification task which facilitates semi-supervised learning. Note that no KL-grade labels are used during the incorporation of additional auxiliary data and only the loss function for leg-side classification is calculated to update the encoder.

4. Experimental setup

4.1. Datasets

We used two publicly available datasets, (i.e., MOST (https://most.ucsf.edu) and OAI (https://nda.nih.gov/oai/)) for the supervised learning phase. Note that both datasets have employed a Synaflexer positioning frame to acquire Posterior-Anterior (PA) bilateral images with a fixed flexion at a 10‘ angle of the X-ray beam. The OAI dataset comprised 4,796 individual participants, whereas the data primarily consists of MRI and X-ray images of nine follow-up checkups over 96 months. Similarly, the MOST dataset is a collection of radiographs from 3,026 individuals. However, it includes four follow-up checks over 84 months. Both datasets also differ in terms of the age group of the participants (i.e., OAI considered participants aged between 45–79 years, while this range was 50–79 for the MOST dataset). Mainly, both datasets are developed while considering the presence of OA or the susceptibility of it developed in the selected participants. Both MOST and OAI are primarily approved by the data acquisition sites and, in particular, by the institutional review board of the University of California, San Francisco. All the participants showed their consent to make the dataset publically available with the condition of anonymity. During the unsupervised learning phase, we considered the Cohort Hip & Cohort Knee (CHECK) dataset [66] which consists of 1002 individuals from the Netherlands having early symptoms of OA of the knee. Primarily, this study is based on population-based observations, whereas a total of 8,363 knee X-rays are used in this research. We utilized five-fold cross-validation where subjects were divided into five groups in our experiment 1.

Table 1
5-fold validation quantitative results of our proposed DC-AAE.

	Acc(%)	Precision(%)	Recall(%)	F1-score(%)
Fold-0	74.52	74.20	82.90	75.34
Fold-1	76.37	74.88	77.72	76.13
Fold-2	74.69	73.37	78.76	74.49
Fold-3	77.36	76.02	76.13	77.22
Fold-4	73.71	72.05	82.94	73.33
Average	75.53	74.10	79.69	75.30

4.2. Model configurations

Fig. 4 illustrates the semi-supervised architecture of our proposed method. First, two encoders having four convolutional layers each are applied to the input for effective KL-grade classification. Next, a pooling layer is added subsequently after each convolutional layer. Each convolutional layer extracts meaningful information from knee joint areas and generates feature maps. Subsequently, the extraction of highly relevant features is performed by the pooling and dense layers by properly reducing the dimensions. Note that we employed the max-pooling layer since it offers much-improved performance during the validation process as compared to average pooling. Both encoders encode lateral and medial parts of the knee image into latent codes z_L and z_M, respectively, which are later concatenated together, thus forming z with a dimension of 16 × 16 × 16. It is noted that the validation set is used for determining the size of the latent code. During the regularization stage, a prior distribution $p(d)$ having multivariate Gaussian distribution with unit standard deviation and zero means is applied to the latent representation (z). In the decoder block, we replaced convolutional layers with transposed convolution layers, keeping the remaining structure the same as the encoders.

Each classifier network, which is fed the latent code z, is made up of four layers: (1) a convolutional layer, (2) a max-pooling layer, (3) dense layers, and (4) a softmax layer. One convolutional layer followed by a max-pooling layer is applied in each classifier which helps to capture the relevant features for classification tasks. To provide a prediction, we finally applied dense layers followed by a softmax layer immediately after the max-pooling layer. We also used a dropout layer with a dropout rate of 0.2 between two dense layers in each classifier to avoid overfitting. The discriminator architecture in DC-AEE is used with similar configurations, i.e., containing one convolutional layer followed by a max-pooling layer, and two dense layers followed by a softmax layer.

We performed step-by-step training on all the models. We first used the CHECK dataset for training the DC-AAE network in an unsupervised manner to initialize the model weights and trained first on all datasets (i.e., OAI, MOST, and CHECK datasets) for legs side identification only. The weights learned in this stage were used to initialize the auto-encoder network for the multitask learning phase with KL-grade classification and leg side identification. Fig. 5 shows the learning curve of our DC-AAE for the supervised phase.

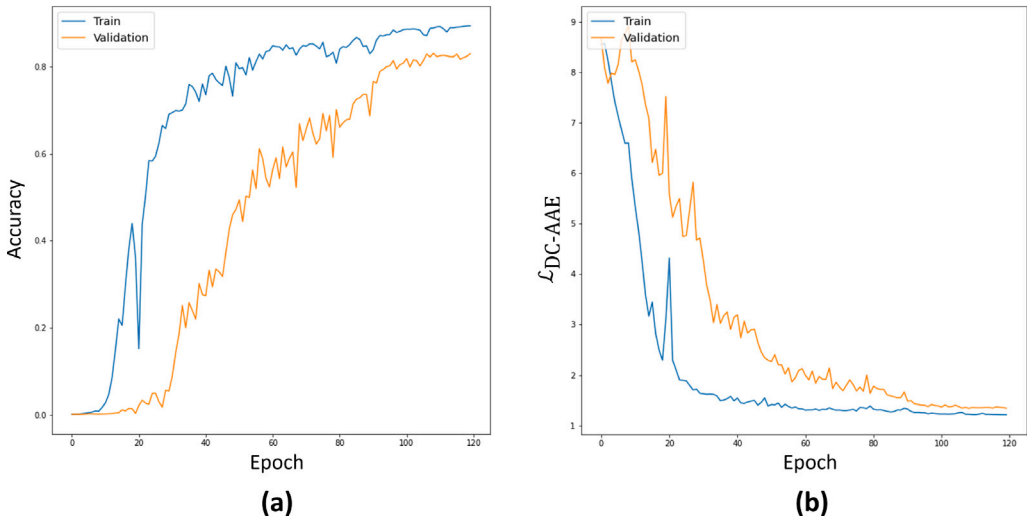

Fig. 5. Learning curves of proposed DC-AAE for KL-grade classification. Figures (a) and (b) show the learning curve for KL-grade classification accuracy and multitask learning loss, respectively.

4.3. Training strategy

The step-by-step training of the proposed model shown in Fig. 4 is performed. We randomly initialize the model and then train the model along an unsupervised path. Specifically, at first, we only train DC-AAE for reconstruction tasks. All the training set X-rays from both OAI and MOST datasets and also additional data from the CHECK dataset are utilized. After optimization of DC-AAE for the reconstruction task, we again used training data to train DC-AAE for the leg side identification task. Finally, we use only labelled data from both datasets (i.e., OAI and MOST) for training DC-AAE with both tasks, i.e., KL-grade classification and leg side identification. Here, we freeze the decoder branch since it has already been optimized. In the supervised phase, networks have been trained for two tasks (i.e., KL-grade, and leg-side classification) KL-grade classification is considered the main task whereas leg-side identification is considered an auxiliary task. Subsequently, more weightage is assigned to the KL-grade classification task during the final training step.

The training was performed with the NVIDIA TESLA V100 TENSOR CORE GPU, whereas the Stochastic Gradient Descent (SGD) optimizer with a 0.0001 learning rate was selected. The batch size was 16, whereas, after each epoch, the accuracy was calculated for validation. If after five epochs, the improvement in terms of validation accuracy was not significant, the learning rate was reduced to half and the model was replaced with the best of five epochs. The minimum value for the learning rate was set to 0.00001. During the training process, to get a stable distribution of activation values, each convolutional layer is supported by a batch normalization layer [67]. Mainly, the non-linearity layer (i.e., rectified linear unit (ReLU)) is used after the batch normalization layer. Note that the ReLU activation was preferred over a hyperbolic tangent and leaky ReLU, mainly due to its enhanced performance. Additionally, to avoid over-fitting, the early stopping strategy has been employed [68]. This technique halts the training process once it detects no change in the validation loss value, thus reducing the chances of overfitting the network on the training data.

5. Results and discussion

We performed various experiments on two publicly available datasets (i.e., OAI and MOST) to evaluate the proposed scheme. The results are analysed based on various aspects, such as the effect of the dual-channel, multitask learning, and the effect of additional data. Finally, we perform cross-dataset validation to evaluate the robustness of our technique. The following subsection reviews our analysis.

Table 2
Comparison of deep learning method for automatic knee OA KL-grade classification.

Methods	Acc (%)	Pr (%)	Re (%)	F1 (%)
CNN model [59]	69.69	68.83	70.11	69.45
Deep Siamese CNN [17]	66.87	68.23	66.83	67.10
ResNet-34 [69]	73.6	73.52	73.69	72.76
ResNet-34 + CBAM [69]	75.15	**75.18**	75.08	74.82
OsteoHRNET [70]	71.74	–	–	–
Dalia et al. [71]	69.8	–	–	–
Our DC-AAE	**75.53**	74.10	**79.69**	**75.30**

5.1. Overall performance

To assess overall performance, we examine the entire pipeline depicted in Fig. 2. We achieved the benchmark performance with the proposed approach by utilizing three datasets (i.e., CHECK, MOST, and OAI) for training the model in the unsupervised phase and supervised phase. The five-fold confusion matrices are shown in Fig. 6(a) to (e) demonstrate the performance of the proposed scheme on OAI datasets. Results show remarkable performance for severe classes (i.e., grades 3 and 4) and slight confusion between earlier grades (1 & 2) has been observed. It is important to note that all the confusion is about the nearby classes, which makes our technique more acceptable from a clinical perspective.

We also compared the performance of the proposed scheme with an existing state-of-the-art technique, without using the CHECK dataset for a fair comparison, which is shown in Table 2. The results demonstrate that the proposed technique outperforms the previous techniques. The reason for such performance is the ability of the proposed architecture to incorporate additional data, which has been exploited in unsupervised as well as supervised phases. The proposed model utilizes the encoding branches, which also play a significant role in improving the learning ability of the network. This helps to extract the most meaningful information, which further eases classifier networks. To demonstrate the effectiveness of the reconstruction task, we have further enhanced our experimentation by removing the decoder part of our proposed DC-AAE using multi-task learning. We referred to this model as model 2 in Table 4 and the results are also summarized in Table 4. The results

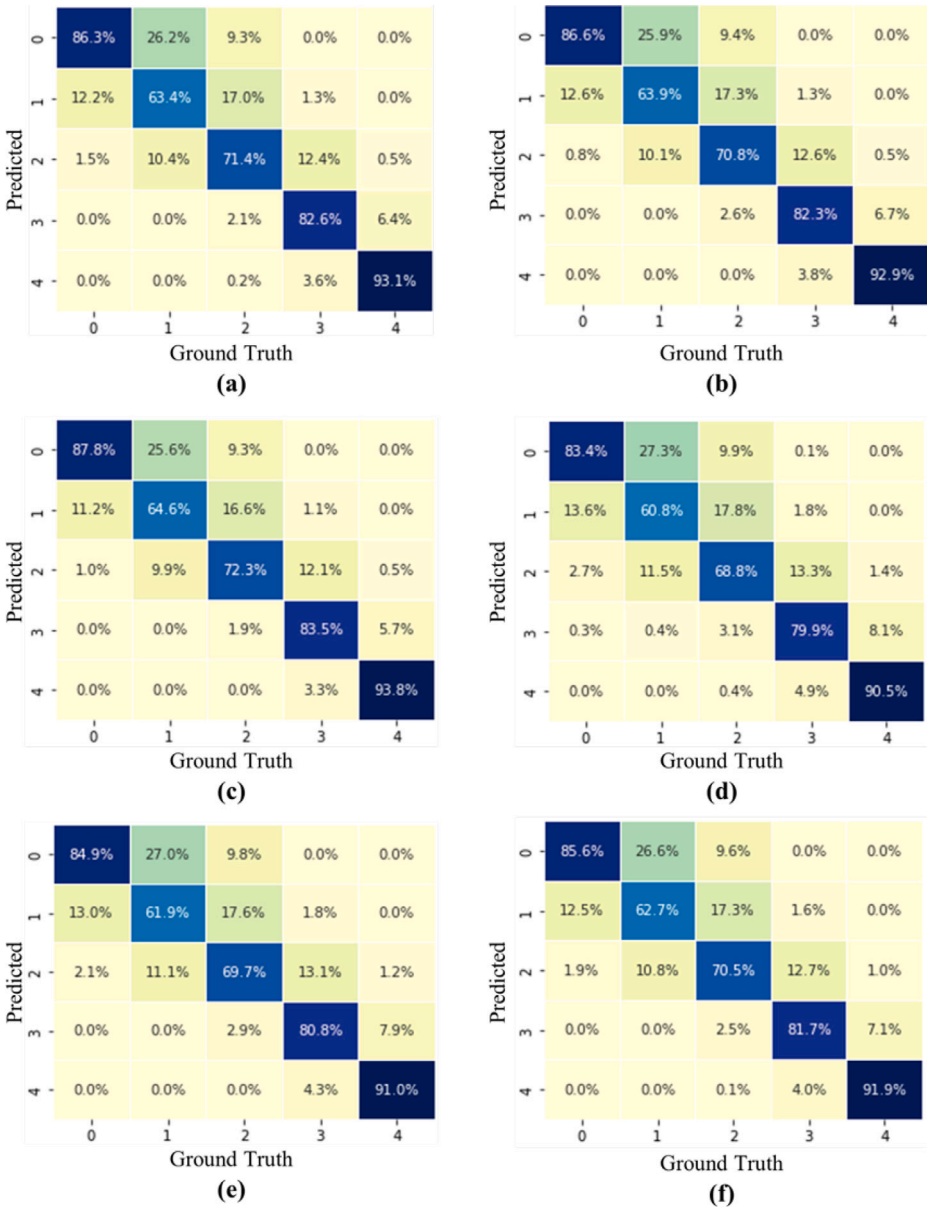

Fig. 6. 5-fold Confusion matrix of the proposed semi-supervised multi-task learning based DC-AAE (a) to (e) five folds of cross-data validation and (f) are average result.

demonstrate that the performance of our proposed model decreased significantly after removing the decoder part. The reason is that the reconstruction task helps the network to learn the most representative latent features. Subsequently, the negative impact of removing the decoder is evident from the degradation of 3.05%, 3.68%, 3.22%, and 3.57% in accuracy, precision, recall, and F1 score, respectively.

5.2. Effect of dual channel autoencoder

In the proposed approach, we employed dual encoding channels to extract meaningful information from the lateral and medial parts of the knee. To investigate the effect of dual encoding channels, we compare the proposed architecture with the conventional single encoder architecture by keeping the remaining part the same as the proposed design and using the whole joint image as an input. We trained the single encoder-based architecture in the same way as we followed for DC-AAE. Fig. 7(a) and (b) present the confusion matrix which contains the results of single encoder-based architecture on MOST and OAI datasets. It can be noticed that the classification error is significantly increased with SE-AE, especially for grade 2 where confusion with adjacent classes is increased. Most importantly, in a few cases, extreme grades (i.e., 1 and 4) are also confused, which is a serious drawback of SE-AAE.

To further investigate the effect of the dual encoding channel, we visualize the activation maps using the Grad-CAM algorithm [72], which represents the network focus regions. Fig. 8(a) and (b) demonstrate the Grad-CAM output of the SE-AAE and DC-AAE architectures, respectively. These results show that the dual encoding channel helps the network to focus on the knee joint region, which contains the most significant information for the estimation of KL-grade. Subsequently, DC-AAE learns the most relevant features which eases the classification of KL-grade.

5.3. Effect of secondary task

To incorporate the unlabelled data in the supervised learning phase, we introduced an additional task, i.e., leg side identification, as an

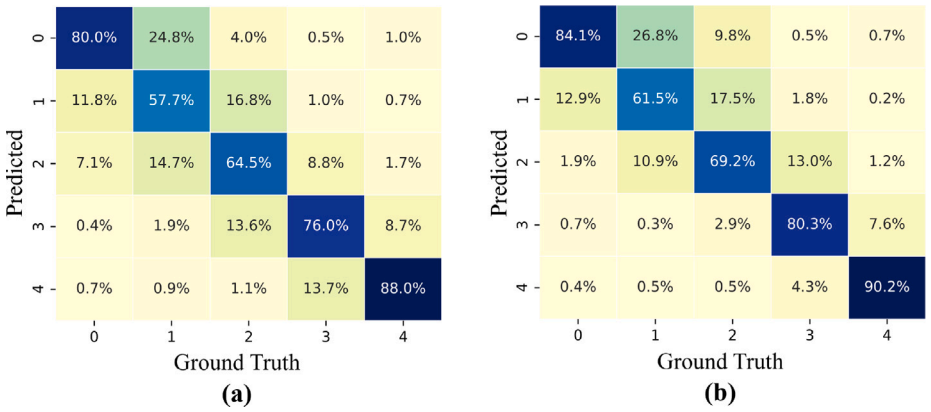

Fig. 7. Confusion matrix of conventional AAE with single encoder (a) on the MOST dataset and (b) on OAI dataset.

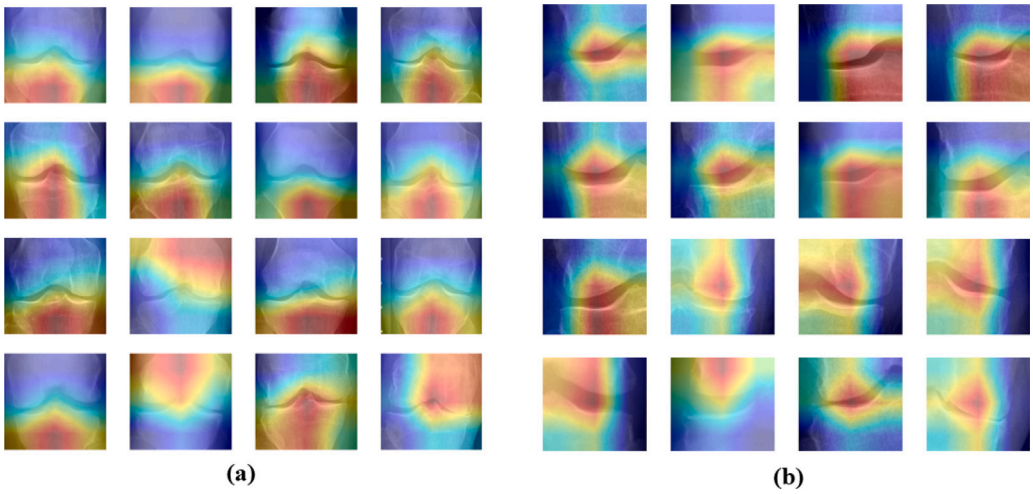

Fig. 8. (a) Grad-CAM output of SE-AE and (b) is our DC-AAE model attention map which correctly focused on the knee joint.

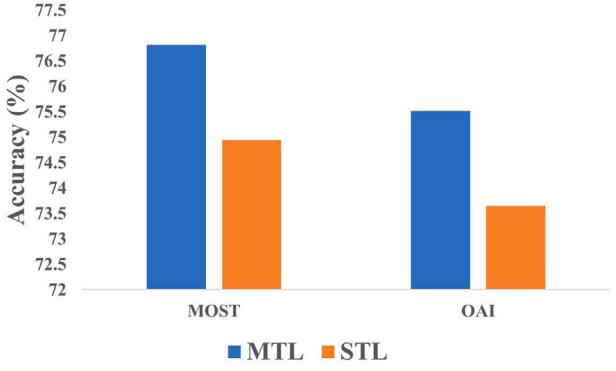

Fig. 9. The results of proposed DC-AAE shown with and without an auxiliary task.

auxiliary task. To analyse the effect of the secondary task, we trained the proposed architecture with one task (i.e., KL-grade classification) in the same settings as we trained our MTL-based model. We first performed weight initialization of DC-AAE in an unsupervised manner with a training set that included three datasets (i.e., OAI, MOST, and CHECK datasets). However, after the unsupervised phase, we trained the model with only KL-grades labels. Fig. 9 presents the results from both techniques (i.e., single task learning (STL) and MTL) on the test set from MOST and OAI datasets. It can be observed that the model with multitasking has achieved better results as compared to STL. The introduction of the additional task enables our DC-AAE networks to exploit the additional data during the supervised learning phase, enhancing the model learning ability and subsequently achieving better results.

5.4. Effect of additional data

We exploit the additional unlabelled dataset, i.e., the CHECK dataset, in both unsupervised and supervised learning phases. To analyse the effect of the additional unlabelled dataset, we perform experimentation without additional data and with different amounts of additional data. We train our DC-AAE architecture with a multitask learning-based approach while considering the KL-grade classification as primary and leg side identification as a secondary task. At first, we do not use any unlabelled data, so only the supervised learning phase is executed. However, we gradually introduce the addition of unlabelled data in different amounts (i.e., 1000, 2000, 4000, and 8,363). Fig. 11 presents the results of these experiments where we observed a slight improvement in the results when we used 1000 additional unlabelled images. However, when we increase the amount of data, the performance also increases. Therefore, summarizing Fig. 11, it can be noted that improvement in the auxiliary tasks while adding additional data eventually helps to improve the performance of the primary task, which cements the contribution of this study in proposing a multi-task semi-supervised framework for KL-grade classification. Intuitively,

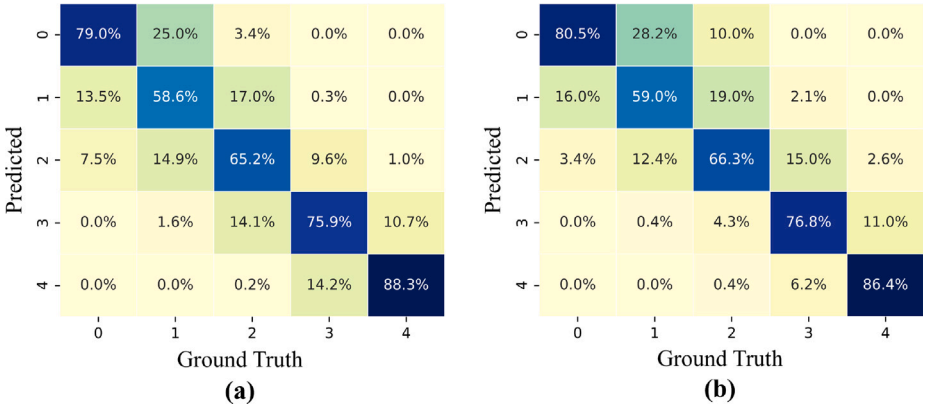

Fig. 10. Results of our DC-AAE using multitask learning model, (a) training on OAI dataset and test on MOST dataset, (b) trained on MOST dataset and test on OAI dataset.

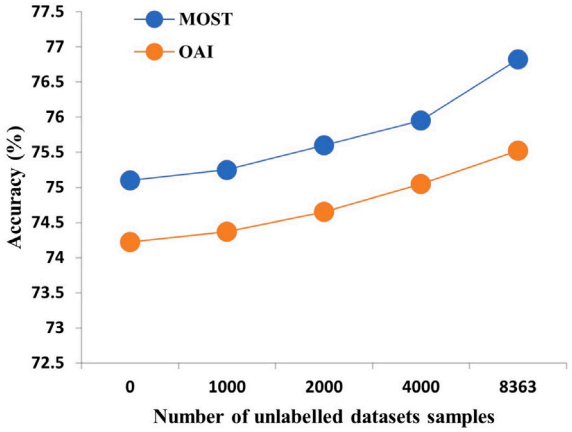

Fig. 11. Illustrates the effect of incorporating the additional unlabelled data on the performance of KL-grade classification on OAI and MOST datasets.

through the feed of additional data for the auxiliary tasks, a better representation of the intrinsic properties of KL-grading is achieved, which eventually improves the performance of KL-grade classification. However, it demonstrates the potential of the proposed architecture to utilize unlabelled data for the improvement of KL-grade classification.

5.5. Robustness analysis

To evaluate the robustness of the proposed scheme, we perform cross-data validation on OAI and MOST datasets. We train the model with two different dataset distributions; specifically, we first train the model on all the OAI and CHECK datasets while following the same training strategy mentioned in Section 4, and then evaluate on the MOST dataset. Similarly, in another setting, we utilize the whole MOST and CHECK datasets for training and testing on OAI datasets. Table 3 compares the performance of cross-dataset experiments with our benchmark performance. It can be noticed that the model shows promising results even on cross-dataset validation; however, performance is slightly degraded due to the reduction of training data in cross-dataset validation. Fig. 10(a) and (b) illustrate the confusion matrix of cross-dataset experimentation on MOST and OAI datasets, respectively. It can be noticed that confusion in all the classes remains in the adjacent classes, and most importantly, there is no confusion between extreme classes (i.e., KL-grade 0 and 4), which depicts the ability of the network to learn meaningful information, even with limited data, that plays a critical role in distinguishing the extreme classes. Furthermore, to assess the diagnostic ability of our proposed model, the receiver operating characteristic (ROC) curves for all five

Table 3
Robustness analysis of our proposed DC-AAE model.

Methods	Acc (%)	Pr (%)	Re (%)	F1 (%)
Our DC-AAE (trained on MOST dataset and tested on OAI dataset)	71.03	69.33	73.82	70.53
Our DC-AAE (trained on OAI dataset and tested on MOST dataset)	73.76	70.79	73.41	71.87

folds (a to e) in Fig. 12 represent each OA severity grade, while the associated area under the curve (AUC) values also shown in Fig. 12 demonstrate the exceptional performance of our model in classifying successive grades. Notably, the classification metrics for grade one have lower values compared to the other grades.

6. Ablation study

We performed an ablation study on key components in the proposed model to quantify the contribution of each component to the overall performance of the model. The following results show that all of them are essential to the proposed model and jointly produce high-quality results in KL-grade classification. By removing each component one by one and the rest of the model remains unchanged, as shown in Table 4 the proposed DC-AAE remarkably improves the performance. After analysing the results, we can learn: (1) The adversarial loss function helps to regularize the encoders and improve their generalization ability, (2) The reconstruction path in unsupervised learning empowers the encoders to extract more meaningful features from the input data, (3) The addition of a new auxiliary task (classification of leg side) enabled us to utilize the additional unlabelled dataset, leading to improved performance of the primary task (classifying knee osteoarthritis severity), (4) we can observe that dual encoder not only improve the model attention to the knee joint area but also improve the model performance. The results indicate that the model utilizing the concatenation of the latent representation outperforms the model using the whole frame as input with a single encoder. Concretely, the dual-encoder-based architecture achieved higher accuracy, precision, recall, and F1 score values of 1.36%, 1.78%, 0.41%, and 1.65%, respectively. This can be attributed to the fact that the use of two cropped inputs with dual-encoder-based architecture helps to focus more on the knee joint area which plays a crucial role to determine the KL-grade.

Furthermore, the mean (standard deviation (SD)) area under the receiver operating characteristic curve (AUC) of five-fold validation was 0.93 ± 0.01 along with standard deviation ± 1 see Fig. 13. Please note, in five-fold cross-validation, an AUC is provided for the model performance in each fold. The mean cross-validated AUC is the average of the AUC values across the five folds. We have incorporated the mean ROC curve with standard deviation plotted as a shaded area. As it can

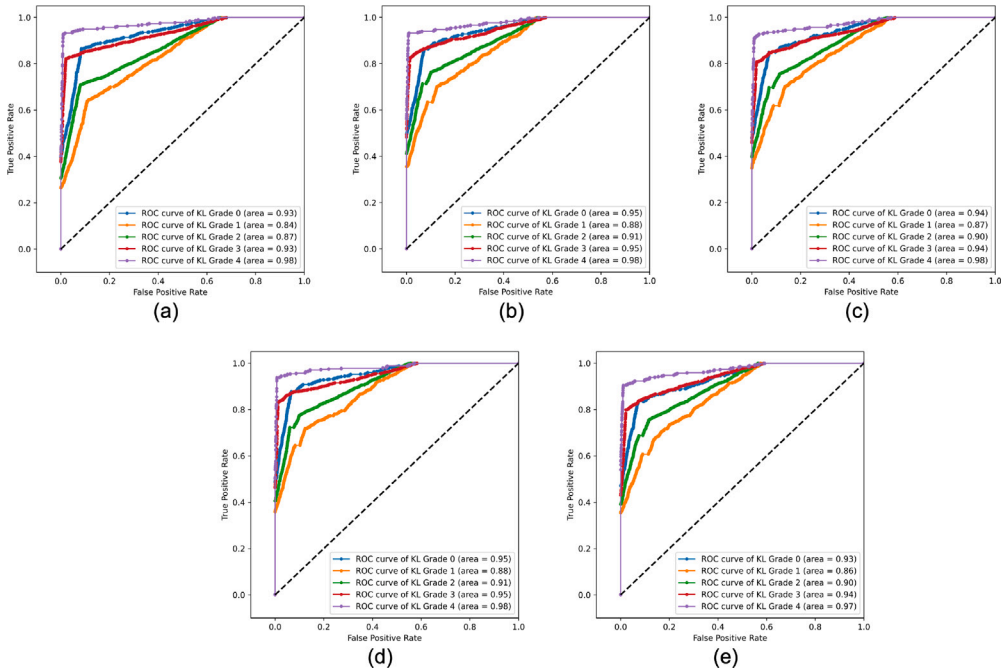

Fig. 12. The illustration of ROC curves obtained on all five folds (i.e., from (a) to (e)). Each subfigure represents each fold and contains the ROC curves for each individual KL-grade (0 4) using one-vs-all (or one-vs-rest) strategy.

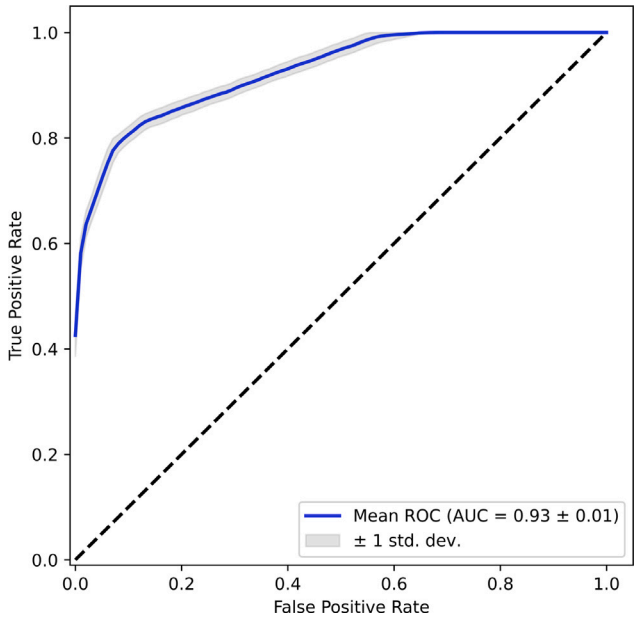

Fig. 13. The illustration of mean ROC curve of five-cross validated folds with standard deviation as shaded area.

be seen that the width of the shaded area is very little which shows that there is no significant difference among the five-folds' performance.

We further delve into the influence of trade-off parameters, α (Eq. (4)) and β (Eq. (5)), which represent the weights of unsupervised tasks, as well as supervised primary and secondary tasks, respectively.

Fig. 14 illustrates the effects of adjusting the weights, α and β. In the preliminary experiments, the β value is kept constant at 0.5 and 1. Specifically, we first assign equal weights to both the primary and auxiliary supervised tasks while altering the weights (ranging from 0.1 to 1.0) assigned to the unsupervised task. Subsequently, in the second experiment, we allocate the highest weight to the primary tasks, zero weight to the auxiliary task, and continue to vary the weights (ranging from 0.1 to 1.0) for the unsupervised task. The outcomes of these initial two experiments are succinctly summarized in Fig. 14(a) and (b). It is discernible that extremely low or high weights of α negatively impact the performance of KL-grade classification. Nevertheless, an α range between 0.4 and 0.7 yielded superior results in both scenarios.

These findings suggest that careful calibration of unsupervised task weights using α can augment system performance. However, an optimal α range must be identified to achieve peak performance.

Subsequent experiments were conducted with the α value fixed at 0.5 and 1, and varying β weights (0.1 to 0.9), which determine the weights for both the primary and auxiliary tasks (as depicted in Eq. (5)). As demonstrated by Fig. 14(c) and (d), smaller β values lead to subpar performance, as the framework predominantly operates as a single-task mechanism. Yet, it is important to note that overemphasis on auxiliary tasks through high β values also deteriorates performance, reaffirming the significance of balancing the auxiliary task's contribution.

7. Conclusion

This study proposes a novel dual-stage end-to-end fully automatic Kellgren-Lawrence (KL) grade classification scheme, which is specially designed for KL-grade classification by using the knee joint regions. Our framework is a semi-supervised multitask learning-based dual-channel adversarial autoencoders (DC-AAE) for KL-grade classification that enables the exploitation of unlabelled data in supervised and unsupervised phases. Therefore, other than the largest publicly available labelled datasets (i.e., OAI and MOST), we also utilize the CHECK dataset without KL-grade labels. The step-by-step training on the proposed DC-AAE architecture is performed. Specifically, we first train the model in an unsupervised manner on all the training sets, incorporating the additional unlabelled CHECK dataset and OAI and MOST datasets. To utilize the same training set in a supervised manner, we adopt multitask learning by introducing an auxiliary task, i.e., leg side identification, along with the primary task of KL-grade classification. The proposed technique has been rigorously evaluated on publicly available datasets (i.e., OAI and MOST) with various aspects such as overall performance, robustness, the effect of additional data, and task. The

Table 4
Ablation study of various configurations of our proposed DC-AAE model.

Model #	Adversarial loss	Reconstruction loss	Multi-task learning		Encoder		Evaluation matrices			
			Leg side classification	KL-grade classification	1st	2nd	Acc	Precision	Recall	F1 score
1	✗	✓	✓	✓	✓	✓	72.09	69.99	74.58	71.26
2	✓	✗	✓	✓	✓	✓	72.48	70.42	75.29	71.73
3	✓	✓	✗	✓	✓	✓	73.00	71.07	75.67	72.35
4	✓	✓	✓	✓	✓	✗	74.17	72.32	77.08	73.65
5	✓	✓	✓	✓	✓	✓	75.53	74.10	78.51	75.30

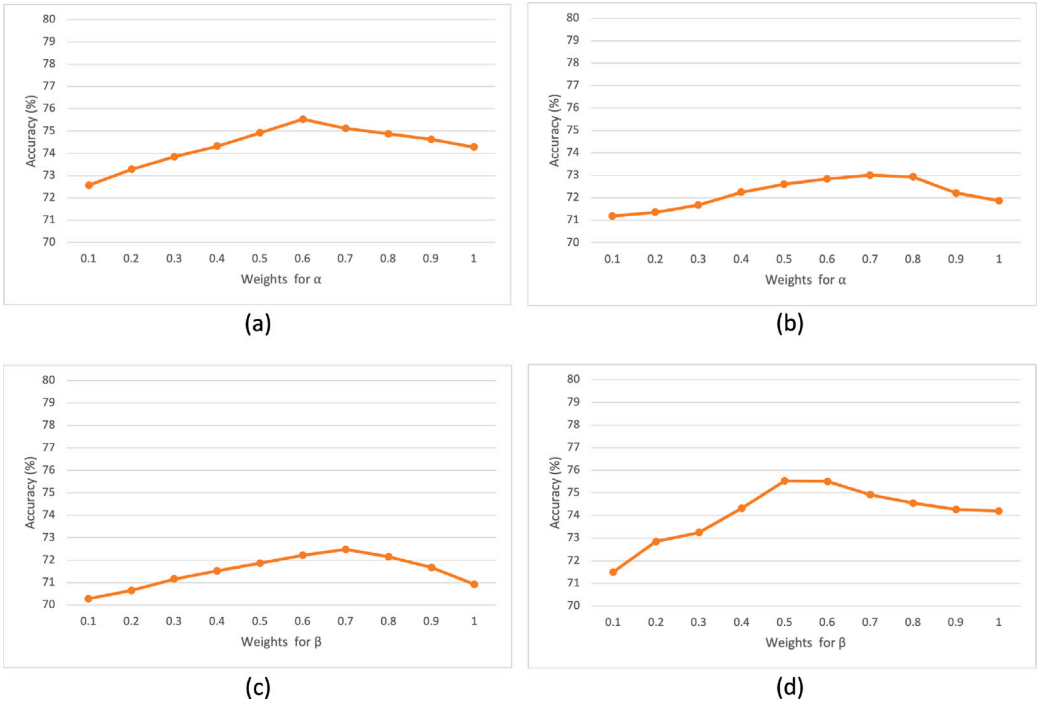

Fig. 14. Illustration of the influence of altering α and β values on KL-grade classification performance. Subfigures (a) and (b) depict the performance effects with a variation of α, holding β constant at 0.5 and 1, respectively. Conversely, subfigures (c) and (d) demonstrate the performance consequences of modulating β values, while maintaining α steady at 0 and 1, respectively.

results demonstrate that the proposed method not only outperformed the state-of-the-art techniques but also showed significant robustness while cross-data analysis. Furthermore, we also investigated the effect of introducing dual encoding channels at the encoding branch and demonstrated its effectiveness by presenting attention maps.

CRediT authorship contribution statement

Muhammad Umar Farooq: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing. **Zahid Ullah:** Methodology, Investigation, Writing – original draft, Writing – review & editing. **Asifullah Khan:** Formal analysis, Investigation, Writing – review & editing. **Jeonghwan Gwak:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), South Korea funded by the Ministry of Education (Grant No. NRF-2020R1I1A307 4141; RS-2023-00248444), and “Regional Innovation Strategy (RIS)” through the National Research Foundation of Korea (NRF), South Korea funded by the Ministry of Education (MOE) (2021RIS-001).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2023.107570.

References

[1] D.A. Varghese, Prevalence of latent tuberculosis in patients with rheumatoid arthritis and ankylosing spondylitis (Ph.D. thesis), Christian Medical College, Vellore, 2020.
[2] P.A. Dieppe, L.S. Lohmander, Pathogenesis and management of pain in osteoarthritis, Lancet 365 (9463) (2005) 965–973.
[3] A. Mobasheri, M. Batt, An update on the pathophysiology of osteoarthritis, Ann. Phys. Rehabil. Med. 59 (5–6) (2016) 333–339.
[4] A. Mathiessen, P.G. Conaghan, Synovitis in osteoarthritis: Current understanding with therapeutic implications, Arthritis Res. Ther. 19 (1) (2017) 1–9.

[5] M. Cross, E. Smith, D. Hoy, S. Nolte, I. Ackerman, M. Fransen, L. Bridgett, S. Williams, F. Guillemin, C.L. Hill, et al., The global burden of hip and knee osteoarthritis: Estimates from the global burden of disease 2010 study, Ann. Rheum. Dis. 73 (7) (2014) 1323–1330.

[6] B. Heidari, Knee osteoarthritis prevalence, risk factors, pathogenesis and features: Part i, Casp. J. Int. Med. 2 (2) (2011) 205.

[7] J.W. Bijlsma, F. Berenbaum, F.P. Lafeber, Osteoarthritis: An update with relevance for clinical practice, Lancet 377 (9783) (2011) 2115–2126.

[8] E. Losina, M.E. Daigle, L. Suter, D. Hunter, D. Solomon, R. Walensky, J. Jordan, S.A. Burbine, A.D. Paltiel, J.N. Katz, Disease-modifying drugs for knee osteoarthritis: Can they be cost-effective? Osteoarthr. Cartil. 21 (5) (2013) 655–667.

[9] K. Lim, C.S. Lau, Perception is everything: OA is exciting, Int. J. Rheum. Dis. 14 (2) (2011) 111–112.

[10] C. Palazzo, C. Nguyen, M.-M. Lefevre-Colau, F. Rannou, S. Poiraudeau, Risk factors and burden of osteoarthritis, Ann. Phys. Rehabil. Med. 59 (3) (2016) 134–138.

[11] S. Salih, A. Hamer, Hip and knee replacement, Surgery (Oxford) 31 (9) (2013) 482–487.

[12] A.J. Carr, O. Robertsson, S. Graves, A.J. Price, N.K. Arden, A. Judge, D.J. Beard, Knee replacement, Lancet 379 (9823) (2012) 1331–1340.

[13] B.S. Ferket, Z. Feldman, J. Zhou, E.H. Oei, S.M. Bierma-Zeinstra, M. Mazumdar, Impact of total knee replacement practice: Cost effectiveness analysis of data from the osteoarthritis initiative, bmj 356 (2017).

[14] P. Baker, K. Muthumayandi, C. Gerrand, B. Kleim, K. Bettinson, D. Deehan, Influence of body mass index (BMI) on functional improvements at 3 years following total knee replacement: A retrospective cohort study, PLoS One 8 (3) (2013) e59079.

[15] K. Baker, T. McAlindon, Exercise for knee osteoarthritis, Curr. Opin. Rheumatol. 12 (5) (2000) 456–463.

[16] J. Kellgren, J. Lawrence, Radiological assessment of osteo-arthrosis, Ann. Rheum. Dis. 16 (4) (1957) 494.

[17] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, S. Saarakkala, Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach, Sci. Rep. 8 (1) (2018) 1–10.

[18] J. Buckland-Wright, I. Carmichael, S. Walker, Quantitative microfocal radiography accurately detects joint changes in rheumatoid arthritis, Ann. Rheum. Dis. 45 (5) (1986) 379–383.

[19] Z. Ullah, M. Usman, S. Latif, A. Khan, J. Gwak, SSMD-UNet: Semi-supervised multi-task decoders network for diabetic retinopathy segmentation, Sci. Rep. 13 (1) (2023) 9087.

[20] Z. Ullah, M. Usman, J. Gwak, MTSS-AAE: Multi-task semi-supervised adversarial autoencoding for COVID-19 detection based on chest X-ray images, Expert Syst. Appl. 216 (2023) 119475.

[21] M. Usman, A. Rehman, A. Shahid, S. Latif, S.S. Byon, S.H. Kim, T.M. Khan, Y.G. Shin, MESAHA-Net: Multi-encoders based self-adaptive hard attention network with maximum intensity projections for lung nodule segmentation in CT scan, 2023, arXiv preprint arXiv:2304.01576.

[22] Z. Ullah, M. Usman, S. Latif, J. Gwak, Densely attention mechanism based network for COVID-19 detection in chest X-rays, Sci. Rep. 13 (1) (2023) 261.

[23] N. Orlov, L. Shamir, T. Macura, J. Johnston, D.M. Eckley, I.G. Goldberg, WND-CHARM: Multi-purpose image classification using compound image transforms, Pattern Recognit. Lett. 29 (11) (2008) 1684–1693.

[24] L. Shamir, N. Orlov, D.M. Eckley, T. Macura, J. Johnston, I. Goldberg, WND-CHARM: Multi-Purpose Image Classifier, Astrophysics Source Code Library, 2013, pp. ascl–1312.

[25] L. Shamir, N. Orlov, D.M. Eckley, T. Macura, J. Johnston, I.G. Goldberg, Wndchrm–an open source utility for biological image analysis, Source Code Biol. Med. 3 (1) (2008) 1–13.

[26] M. Usman, B.-D. Lee, S.-S. Byon, S.-H. Kim, B.-i. Lee, Y.-G. Shin, Volumetric lung nodule segmentation using adaptive roi with multi-view residual learning, Sci. Rep. 10 (1) (2020) 12839.

[27] A. Rehman, M. Usman, A. Shahid, S. Latif, J. Qadir, Selective deeply supervised multi-scale attention network for brain tumor segmentation, Sensors 23 (4) (2023) 2346.

[28] M. Usman, Y.-G. Shin, DEHA-Net: A dual-encoder-based hard attention network with an adaptive ROI mechanism for lung nodule segmentation, Sensors 23 (4) (2023) 1989.

[29] D. Dai, C. Dong, S. Xu, Q. Yan, Z. Li, C. Zhang, N. Luo, Ms RED: A novel multi-scale residual encoding and decoding network for skin lesion segmentation, Med. Image Anal. 75 (2022) 102293.

[30] M. Usman, S. Latif, M. Asim, B.-D. Lee, J. Qadir, Retrospective motion correction in multishot MRI using generative adversarial network, Sci. Rep. 10 (1) (2020) 4786.

[31] S. Latif, M. Asim, M. Usman, J. Qadir, R. Rana, Automating motion correction in multishot MRI using generative adversarial networks, 2018, arXiv preprint arXiv:1811.09750.

[32] M. Usman, A. Rehman, A. Shahid, S. Latif, S.S. Byon, B.D. Lee, S.H. Kim, Y.G. Shin, et al., MEDS-Net: Self-distilled multi-encoders network with bi-direction maximum intensity projections for lung nodule detection, 2022, arXiv preprint arXiv:2211.00003.

[33] J. Antony, Automatic Quantification of Radiographic Knee Osteoarthritis Severity and Associated Diagnostic Features Using Deep Convolutional Neural Networks (Ph.D. thesis), Dublin City University, 2018.

[34] J. Antony, K. McGuinness, K. Moran, N.E. O'Connor, 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, 2017, pp. 376–390.

[35] S. Laine, T. Aila, Temporal ensembling for semi-supervised learning, 2016, arXiv preprint arXiv:1610.02242.

[36] V. Verma, K. Kawaguchi, A. Lamb, J. Kannala, Y. Bengio, D. Lopez-Paz, Interpolation consistency training for semi-supervised learning, 2019, arXiv preprint arXiv:1903.03825.

[37] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, C. Raffel, Mixmatch: A holistic approach to semi-supervised learning, 2019, arXiv preprint arXiv:1905.02249.

[38] R. Caruana, Multitask learning, Mach. Learn. 28 (1) (1997) 41–75.

[39] S. Latif, R. Rana, S. Khalifa, R. Jurdak, J. Epps, B.W. Schuller, Multi-task semi-supervised adversarial autoencoding for speech emotion recognition, IEEE Trans. Affect. Comput. (2020).

[40] A. Amyar, R. Modzelewski, H. Li, S. Ruan, Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation, Comput. Biol. Med. 126 (2020) 104037.

[41] Y. Zhang, H. Li, J. Du, J. Qin, T. Wang, Y. Chen, B. Liu, W. Gao, G. Ma, B. Lei, 3D multi-attention guided multi-task learning network for automatic gastric tumor segmentation and lymph node classification, IEEE Trans. Med. Imaging 40 (6) (2021) 1618–1631.

[42] S. Sahu, R. Gupta, G. Sivaraman, W. AbdAlmageed, C. Espy-Wilson, Adversarial auto-encoders for speech based emotion recognition, 2018, arXiv preprint arXiv:1806.02146.

[43] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, B. Frey, Adversarial autoencoders, 2015, arXiv preprint arXiv:1511.05644.

[44] H. Oka, S. Muraki, T. Akune, A. Mabuchi, T. Suzuki, H. Yoshida, S. Yamamoto, K. Nakamura, N. Yoshimura, H. Kawaguchi, Fully automatic quantification of knee osteoarthritis severity on plain radiographs, Osteoarthr. Cartil. 16 (11) (2008) 1300–1306.

[45] L. Shamir, S.M. Ling, W. Scott, M. Hochberg, L. Ferrucci, I.G. Goldberg, Early detection of radiographic knee osteoarthritis using computer-aided analysis, Osteoarthr. Cartil. 17 (10) (2009) 1307–1312.

[46] L. Shamir, S.M. Ling, W.W. Scott, A. Bos, N. Orlov, T.J. Macura, D.M. Eckley, L. Ferrucci, I.G. Goldberg, Knee x-ray image analysis method for automated detection of osteoarthritis, IEEE Trans. Biomed. Eng. 56 (2) (2008) 407–415.

[47] D. Saini, T. Chand, D.K. Chouhan, M. Prakash, A comparative analysis of automatic classification and grading methods for knee osteoarthritis focussing on X-ray images, Biocybern. Biomed. Eng. (2021).

[48] A.Q. Zafar, R. Zamani, M. Akrami, The effectiveness of foot orthoses in the treatment of medial knee osteoarthritis: A systematic review, Gait Posture 76 (2020) 238–251.

[49] S. Moustakidis, N.I. Papandrianos, E. Christodolou, E. Papageorgiou, D. Tsaopoulos, Dense neural networks in knee osteoarthritis classification: A study on accuracy and fairness, Neural Comput. Appl. (2020) 1–13.

[50] T.K. Yoo, D.W. Kim, S.B. Choi, E. Oh, J.S. Park, Simple scoring system and artificial neural network for knee osteoarthritis risk prediction: A cross-sectional study, PLoS One 11 (2) (2016) e0148724.

[51] A. Khan, A. Sohail, U. Zahoora, A.S. Qureshi, A survey of the recent architectures of deep convolutional neural networks, Artif. Intell. Rev. 53 (8) (2020) 5455–5516.

[52] S.-E. Moon, C.-J. Chen, C.-J. Hsieh, J.-L. Wang, J.-S. Lee, Emotional EEG classification using connectivity features and convolutional neural networks, Neural Netw. 132 (2020) 96–107.

[53] G. Lou, H. Shi, Face image recognition based on convolutional neural network, China Commun. 17 (2) (2020) 117–124.

[54] K.N. Sukhia, M.M. Riaz, A. Ghafoor, S.S. Ali, Content-based remote sensing image retrieval using multi-scale local ternary pattern, Digit. Signal Process. 104 (2020) 102765.

[55] H. Ramadan, C. Lachqar, H. Tairi, Saliency-guided automatic detection and segmentation of tumor in breast ultrasound images, Biomed. Signal Process. Control 60 (2020) 101945.

[56] K.C. Kim, H.C. Cho, T.J. Jang, J.M. Choi, J.K. Seo, Automatic detection and segmentation of lumbar vertebrae from X-ray images for compression fracture evaluation, Comput. Methods Programs Biomed. 200 (2021) 105833.

[57] S. Yang, Feature Engineering in Fine-Grained Image Classification (Ph.D. thesis), 2013.

[58] H.H. Nguyen, S. Saarakkala, M.B. Blaschko, A. Tiulpin, Semixup: In-and out-of-manifold regularization for deep semi-supervised knee osteoarthritis severity grading from plain radiographs, IEEE Trans. Med. Imaging 39 (12) (2020) 4346–4356.

[59] P. Chen, L. Gao, X. Shi, K. Allen, L. Yang, Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss, Comput. Med. Imaging Graph. 75 (2019) 84–92.

[60] M. Górriz, J. Antony, K. McGuinness, X. Giró-i Nieto, N.E. O'Connor, Assessing knee OA severity with CNN attention-based end-to-end architectures, in: International Conference on Medical Imaging with Deep Learning, PMLR, 2019, pp. 197–214.

[61] A. Swiecicki, N. Li, J. O'Donnell, N. Said, J. Yang, R.C. Mather, W.A. Jiranek, M.A. Mazurowski, Deep learning-based algorithm for assessment of knee osteoarthritis severity in radiographs matches performance of radiologists, Comput. Biol. Med. 133 (2021) 104334.

[62] Y. Nasser, R. Jennane, A. Chetouani, E. Lespessailles, M. El Hassouni, Discriminative regularized auto-encoder for early detection of knee OsteoArthritis: Data from the osteoarthritis initiative, IEEE Trans. Med. Imaging 39 (9) (2020) 2976–2984.

[63] J. Antony, K. McGuinness, N.E. O'Connor, K. Moran, Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks, in: 2016 23rd International Conference on Pattern Recognition, ICPR, IEEE, 2016, pp. 1195–1200.

[64] A. Tiulpin, J. Thevenot, E. Rahtu, S. Saarakkala, A novel method for automatic localization of joint area on knee plain radiographs, in: Scandinavian Conference on Image Analysis, Springer, 2017, pp. 290–301.

[65] Softneta, DICOM library anonymize, share, view DICOM files ONLINE dicomlibrary.com, 2023, https://www.dicomlibrary.com/dicom/dicom-tags/. (Accessed 28 Jan 2023).

[66] J. Wesseling, J. Dekker, W. Van den Berg, S. Bierma-Zeinstra, M. Boers, H. Cats, P. Deckers, K. Gorter, P. Heuts, W. Hilberdink, et al., CHECK (cohort hip and cohort knee): Similarities and differences with the osteoarthritis initiative, Ann. Rheum. Dis. 68 (9) (2009) 1413–1419.

[67] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, PMLR, 2015, pp. 448–456.

[68] L. Prechelt, Early stopping-but when? in: Neural Networks: Tricks of the Trade, Springer, 1998, pp. 55–69.

[69] B. Zhang, J. Tan, K. Cho, G. Chang, C.M. Deniz, Attention-based cnn for kl grade classification: Data from the osteoarthritis initiative, in: 2020 IEEE 17th International Symposium on Biomedical Imaging, ISBI, IEEE, 2020, pp. 731–735.

[70] R.K. Jain, P.K. Sharma, S. Gaj, A. Sur, P. Ghosh, Knee osteoarthritis severity prediction using an attentive multi-scale deep convolutional neural network, 2021, arXiv preprint arXiv:2106.14292.

[71] Y. Dalia, A. Bharath, V. Mayya, S.S. Kamath, Deepoa: Clinical decision support system for early detection and severity grading of knee osteoarthritis, in: 2021 5th International Conference on Computer, Communication and Signal Processing, ICCCSP, IEEE, 2021, pp. 250–255.

[72] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.