



A fully automatic fine tuned deep learning model for knee osteoarthritis detection and progression analysis



Sameh Abd El-Ghany^{a,b}, Mohammed Elmogy^c, A. A. Abd El-Aziz^{a,d,*}

^a Dept. of Information Systems, College of Computer and Information Sciences, Jouf University, Al-Jouf, Saudi Arabia

^b Information Systems Department, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt

^c Information Technology Department, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt

^d Dept. of Information Systems and Technology, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt

ARTICLE INFO

Article history:

Received 13 December 2022

Revised 19 March 2023

Accepted 27 March 2023

Available online 31 March 2023

Keywords:

Knee osteoarthritis

Deep learning

DenseNet169

Fine-tuning

ABSTRACT

Knee osteoarthritis (KOA) sufferers have one of the highest disability-adjusted life years. The entire knee joint is affected by KOA. KOA is a condition that makes it hard for the knee to move normally. In KOA, the damage to the joints is irreversible, and the only treatment is a total knee replacement (TKR), which is expensive and only lasts a short time, especially for obese people. The individual's social isolation and low quality of life are significant outcomes of KOA. Despite being time-consuming and highly subject to user variation, segmentation, manual diagnosis, and annotation of knee joints are still the most common procedure used in clinical practices to diagnose osteoarthritis. To overcome the above limitations of the previously mentioned widely used procedure and reduce diagnostic errors made by doctors, we proposed a fine-tuning KOA diagnosis model using the DenseNet169 deep learning (DL) technique to improve the efficiency of KOA diagnosis. Our model will reduce the cost of diagnosis, speed up diagnosis, and delay disease progression, enhancing the procedure from the patient's perspective. The proposed model will determine the degree of KOA diseases by making multi-classification and binary classifications of the KOA severity. It will successfully localize the opacities' peripheral, diffuse distribution, and vascular thickening. Therefore, the proposed model would enable clinicians to understand the primary causes of KOA through this localization. We evaluated the proposed model over the OAI dataset. The OAI dataset was pre-processed by artifact removal, resizing, contrast handling, and a normalization technique. The proposed model was evaluated and compared with recent classifiers. In multi-classification, the DenseNet169 model achieved 95.93%, 88.77%, 95.41%, 85.8%, and 87.08% for accuracy, sensitivity, specificity, precision, and F1-score, respectively. In binary classification, the accuracy, sensitivity, specificity, precision, and the F1-score of the DenseNet169 model were 93.78%, 91.29%, 91.29%, 87.57%, and 89.27%, respectively. Therefore, the proposed model is an unrivaled perceptive outcome with tuning as opposed to other ongoing existing frameworks.

© 2023 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Computers and Artificial Intelligence, Cairo University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

As the world's population ages, osteoarthritis (OA) is a degenerative condition that worsens. OA is the most prevalent musculoskeletal condition, and it's the most prevalent form of arthritis that results in significant disability for patients worldwide. OA is the leading cause of frailty in overweight and older people. It is a disease of the joints that usually affects the ligament, which could be more than just the cartilage. Cartilage is a smooth, elastic tissue that prevents the bones from rubbing, stabilizes the joint, and

makes it easier for the bones to move. OA is caused by the rupture of protective cartilage, which causes the bones to rub against one another, resulting in excessive joint pain and stiffness. OA most frequently affects the knee, followed by the hip and hand joints. Symptom evaluation and plain radiograph evaluation are used to diagnose OA, but this method is subject to error.

Knee osteoarthritis (KOA) is a widespread condition that ranks as the fourth most common cause of disability among older people and the young worldwide [1,2]. Up to 14 million people in the United States are thought to have symptomatic knees, according to the national health interview survey (NHIS) [3], and tens of millions more are affected in Europe, South America, Asia, or the Middle East [4]. The economic burden of KOA is estimated to amount to

* Corresponding author at: Dept. of Information Systems, College of Computer and Information Sciences, Jouf University, Al-Jouf, Saudi Arabia.

up to 2.5% of the growth of the national product in western nations because of lost productivity and healthcare costs [5]. According to recent research, the global aging population and at least 130 million people will be affected by KOA [6]. More than 20 billion dollars have been spent on health care because of KOA treatment [7].

The Kellgren-Lawrence (KL) [8] scale is used to grade the KOA disease's severity. It has five severity grades ranging from 0 to 4 to quantify KOA. Grade 0 indicates that there are no radiographic KOA features. Grade one suggests doubtful joint space narrowing (JSN) and osteophytic lipping. On an anteroposterior weight-bearing radiograph, grade two indicates the presence of certain osteophytes and the possibility of JSN. Grade three suggests multiple osteophytes, a confirmed JSN, sclerosis, and a possible bony deformity. Finally, the grade four diagnosis means significant bony deformity, large osteophytes, marked JSN, and severe sclerosis [9].

For late-stage KOA, there are few treatment options. Therefore, detecting and evaluating KOA as soon as possible is essential to lessen the patient's suffering. Currently, imaging is used to diagnose KOA and determine its severity. X-ray images, ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) are all common image modalities [10]. The X-ray is the most convenient, widely available, and cost-effective of the four modalities listed above and is responsible for a substantial portion of clinical practice.

X-rays measure the joint space width (JSW), a measure of the distance between the tibia and the femur, which is regarded as a sign of cartilage thickness. The JSN can be defined as a change in JSW over time using individual patient X-rays taken at various time points. Despite these benefits, X-ray is known to be insensitive when attempting to identify early KOA changes due to two factors [11]:

- The lesion only covers a small portion of the X-ray image. The cartilage status is overwhelmed by irrelevant components like clothing, tissues, or muscles, causing final decisions to be erroneous.
- It is difficult to establish standard diagnostic criteria due to the variety of bone shapes and densities. Since the computer-aided system cannot easily incorporate the well-trained radiologist's personal experiences, only personal experiences are used to determine the severity of KOA. Hence, early diagnosis and treatment remain the major obstacles. So, we strive to improve the early diagnosis and treatment of KOA patients using Deep learning (DL).

An emerging field of medical research is artificial intelligence (AI), or computer programs that can learn and reason like humans. DL is a subfield of AI, and it refers to algorithms that can learn independently and generally improve with experience. Utilizing DL in clinical application tools makes it possible to use a broader range of input information to optimize patient care decisions that require a lot of time from a clinician. By offering a suggested course of action and assisting in eliminating clinician bias, the models may also aid in streamlining the process of making clinical decisions, such as in KOA.

In this research, we used DenseNet169 to propose a model that will help clinicians to diagnose KOA disease efficiently and compared DenseNet169's results with the common five DL approaches: InceptionV3, Xception, ResNet50, DenseNet121, and InceptionResNetV2. The proposed model has tremendous potential by easing the burden on primary radiologists and enhancing early diagnosis and treatment. The proposed model would determine the degree of KOA diseases by making multi-classification and binary classifications of the KOA severity. It would successfully localize the opacities' peripheral, diffuse distribution, and vascular thickening. Therefore, the proposed model may enable clinicians to understand the primary causes of KOA through this localization. The following is a summary of the contributions made in this research:

- Utilizing the DenseNet169 as a fully convolutional network (FCN), we proposed an extremely accurate model that automatically locates the KOA in the X-ray images.
- Utilizing the DenseNet169 for a highly accurate multi-classification and binary classification compared to previous methods for determining the severity of KOA.
- The proposed model can diagnose KOA with high accuracy and fast diagnosis of disease, reducing the time and the cost required to diagnose KOA disease and helping clinical to provide a suitable treatment for KOA patients.
- The proposed approach trains the DenseNet169 with a weighted ratio of two loss functions: categorical cross entropy and means squared error, with the natural benefit of predicting KOA severity in ordinal (0, 1, 2, 3, and 4) and continuous (0–4) scales.
- Developing an end-to-end pipeline that automatically combines the Grad-CAM for localizing KOA and the DenseNet169 for quantifying the localized KOA into a system automatically diagnoses KOA.

The remainder of this paper is structured as follows. [Section 2](#) presents a KOA diagnosis system-related literature review. A comprehensive description of the model's materials and architecture can be found in [Section 3](#). [Section 4](#) presents and discusses the implementation and evaluation, while [Section 5](#) provides a conclusion and recommendations for future work.

2. Literature review

Schiratti et al. [12] used a DL technique with 9280 knee MR images from 3268 patients in the osteoarthritis initiative (OAI) database to predict further cartilage degradation as measured by joint space narrowing at 12 months based on MR images and clinical variables like body mass index (BMI). The area under the curve (AUC) for the classification model was 65%, and an AUC score of 58.7% was achieved by trained radiologists on a comparable task, indicating the difficulty of the classification task. Wang et al. [13] proposed a convolutional neural network (CNN) based automatic method for OA diagnosis using the OAI database. The proposed model accuracy was 69.18%. Kondal et al. [14] presented an approach using CNN to grade knee radiographs on the KL scale automatically. There were two interconnected steps in the proposed model: an object detection model separated individual knees from the rest of the image in the first stage, and a regression model automatically assigned a KL scale to each knee in the second stage. The authors demonstrated that the mean absolute error was reduced from 1.09 to 0.28 by fine-tuning the model before testing it on a private hospital dataset. Tiulpin et al. [15] presented an automated KL grading scale-based transparent computer-aided diagnosis method for knee OA. It was based on the deep siamese convolutional neural network. The proposed method was validated on 5,960 knees from the OAI and trained solely with data obtained from the multicenter osteoarthritis study (MOST). Compared to the annotations a committee of clinical experts provided, the proposed method produced an average multi-class accuracy of 66.71% and a quadratic Kappa coefficient of 83%. A radiological OA diagnosis area under the ROC curve of 93% was also reported.

Kokkotis et al. [16] proposed a fuzzy logic-based feature selection as the foundation for the proposed approach, followed by learning algorithms and nan-explainability analysis. While fuzzy logic was utilized to combine multiple feature importance scores, resulting in a more robust selection of informative features, the proposed method aggregated the results of filter, wrapper, and embedded feature selection algorithms. The findings demonstrated that the proposed method could select a subset of risk factors that improve the accuracy of various ML models' performance com-

pared to popular feature selection algorithms. On a group of 21 selected risk factors, the best-performing model of the random forest (RF) classifier achieved a classification accuracy of 73.55%. McCabe et al. [17] proposed two validated models for treating radiological knee osteoarthritis. A diagnostic and prognostic model of time until the onset of KOA was included. The OAI provided model development and optimization data, and the MOST offered external validation for both models. AIC, determined from the test data, was used to optimize logistic regression (LR) and Cox regression. The diagnostic model had an AUC of 67% for the validation data and 75% for the test data.

Liu et al. [18] utilized the region proposal network (RPN) and Fast R-CNN. The RPN was trained to generate region proposals that included the knee joint and were then used for classification by Fast R-CNN. Using CNNs for localized classification, the authors filtered irrelevant information from X-ray images and extracted clinically relevant features. They employed a novel loss function whose weighting scheme permitted to address the class imbalance for further performance enhancements. Additionally, when increasing the input size of X-ray images, more prominent anchors were utilized to address the issue of anchors that did not match the object. The Faster R-CNN achieved sensitivity above 78%, specificity above 94%, and a mean average precision of nearly 0.82. Lim et al. [19] utilized patient statistical data on medical use and CNN to detect the occurrence of osteoarthritis. Five thousand seven hundred forty-nine individuals served as the study's base. From the simple background medical records of the patients, principal component analysis (PCA) with quantile transformer scaling was used to generate features and identify the presence of osteoarthritis. The tests demonstrated that the proposed approach achieved a 76.8% AUC. Tiulpin et al. [20] used raw radiographic data, a patient's medical history, anthropometric data, and optionally a radiologist's statement (KL-grade) to propose a novel ML-based approach to predict structural OA progression. For patients without early or moderate OA, the goal was to anticipate any rise in current KL grade or potential need for TKR within the next seven years of the baseline exam. The authors used a separate, large dataset to test the proposed model. As an auxiliary outcome, a CNN was used in the proposed approach to assessing the likelihood of OA progression alongside the knee's current OA severity. A gradient boosting machine (GBM) combined CNN's prediction with clinical data to enhance the prognosis. The proposed model had a precision of 62% and an AUC of 82%. Tolpadi et al. [21] detected OA lesions in both knees using the KL scale. They proposed a semi-automatic CADx model based on deep siamese CNN and a refined ResNet-34. The average multi-class accuracy of the proposed model was 61%, making it more accurate than KL-1 and KL-2 at classifying KL-0, KL-3, and KL-4.

From the previous review of the current studies conducted recently, we can conclude that no study achieved more than 87% accuracy. However, with the OAI dataset, in multi-classification, the DenseNet169 model achieved 95.93%, 88.77%, 95.41%, 85.8%, and 87.08%, for accuracy, sensitivity, specificity, precision, and f1-score, respectively. In binary classification, the DenseNet169 model achieved 93.78%, 91.29%, 91.29%, 87.57%, and 89.27% for accuracy, sensitivity, specificity, precision, and F1-score, respectively.

3. Materials and methods

3.1. Dataset description

The OAI dataset's baseline cohort includes X-ray and MRI images of 4,746 participants. Based on the assessments made by the Boston University X-ray reading center (BU), a total of 4,446 X-ray images from the entire cohort were chosen. This selection

is based on the availability of KL grades for both knees, and there are 8,892 knee images altogether. The KOA samples are depicted in Fig. 1 according to the KL grades. We divided the OAI datasets into three datasets: the training dataset with 6224 images (70%), the test dataset with 1778 images (20%), and the validation dataset with 889 images (10%). The training dataset was classified into three classes for multi-classification to avoid data leakage, as shown in Table 1. The training dataset was classified into negative and positive classes for binary classification. The negative class means that the image has no KOA disease, and the positive class indicates that the image has a KOA disease.

3.2. Model architecture and training

Currently, the DenseNet169 demonstrates a true classification success. The in-depth proposed model aims to create an automated system to process X-ray images to detect KOA disease. The proposed model will be discussed in the following subsections. Fig. 2 provides the steps of the proposed model: (1) image pre-processing, (2) dataset splitting, (3) pre-training the DenseNet169 model, (4) transfer learning, (5) fine-tuning for the DenseNet169 model, (6) model validation, (7) KOA severity grading, (8) KOA severity localization, (9) using the Grad-CAM algorithm, and (10) KOA disease feedback and diagnosis.

Collecting images was the first step of the proposed model, followed by image pre-processing. Because efficiently pre-processing the collected data leads to accurate results, data pre-processing was occasionally one of the most crucial steps. The pre-processing of OAI images was done by artifact removal, resizing, contrast handling, and normalization. The second step was dividing the OAI dataset into training (70%), testing (20%), and validation (10%) datasets. The ImageNet dataset [22], one of the medical field's most widely used large datasets, was used in the third step to pre-train the DenseNet169. During feature extraction, the pre-trained DenseNet169 model can extract new features from the dataset. The fourth step was applying the transfer learning on the pre-trained DenseNet169 model. Transfer learning is one of the

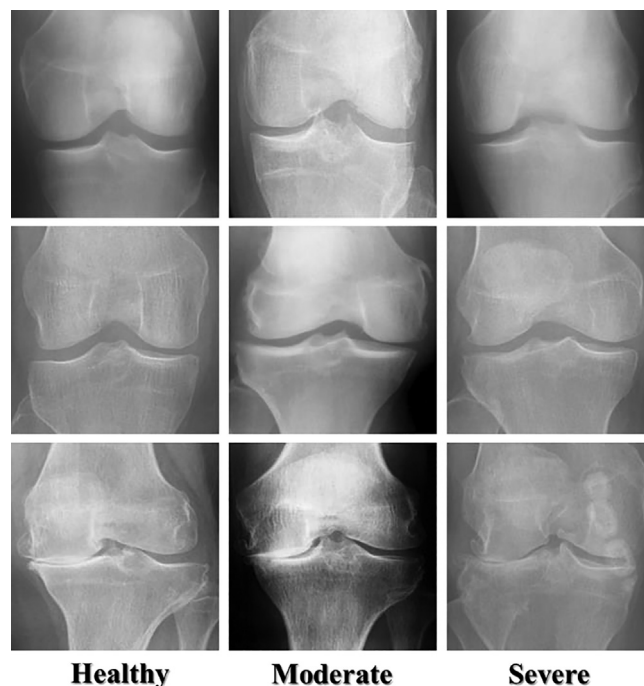


Fig. 1. Samples of knee images.

Table 1

The three classes of the training dataset in multi-classification.

Class	Image Count
Healthy	5294
Moderate	757
Severe	173

most widely used techniques for identifying and classifying KOA cases. In transfer learning, using a limited dataset and feature learning from the DenseNet169 model, related problems that cannot be implemented from scratch were solved [23].

Transfer learning focuses on how to apply one mission's expertise or knowledge to subsequent tasks. It is usually done for undertakings where your dataset has too little information to prepare a full-scale model without any preparation. In the fifth step, to get the best performance of the model in the newly applied domain, a process of parameter tuning that involves updating and rebuilding the model architecture was implemented during the training phase to obtain the optimal network parameters. In fine-tuning step, we unfreeze the model obtained above and retrain it with a very low learning rate on the new data. Gradually adopting the pre-trained features to the new data may result in significant enhancements. The model was tested and validated on the remaining invisible images (the testing dataset). Finally, the DenseNet169 model was deployed after several experiments to determine the KOA severity grading and localization.

3.2.1. Data pre-processing

Since the images vary in size, image resolution, pixel-level noise, symbols, bright text, etc., the data pre-processing step in this research was necessary. An image mask, created by binary thresholding [24], as shown in Eq. (1), was implemented over the images to address such artifacts. In addition, the contrast of X-ray images might change. The training images' contrasts were normalized in the training phase to solve the mentioned problem. After that, we filtered the images to remove noise. Each pixel image was precisely subtracted from the three main colors' average: red, green, and blue (RGB).

$$Mask(m, n) = \begin{cases} \max_i, i(m, n) \geq \min_i \\ 0 \text{ otherwise.} \end{cases} \quad (1)$$

It is possible to define the effect of various intensities of pixels by normalizing data using multiple methods. The normalization approach is used to get the normalized data PI^* from the pixel intensity noted as PI . For each of the three primary colors, this pixel ranges from 0 to 255. As a result, each pixel value was normalized by dividing it by 255. The experiment's maximum and minimum values served as the basis for the normalization, as shown in Eq. (2). The images were resized to a fixed resolution of 224 by 224.

$$P_l^* = (P_{ll} - Min_{old}) \frac{Max_{new} - Min_{new}}{Max_{old} - Min_{old}} + Min_{new}, l \in [0, n] \quad (2)$$

3.2.2. DenseNet169

The architecture of the DenseNet169 comprises two main parts: a classifier and feature extraction. The two essential layers of the DenseNet169 are the convolution and pooling layers. Each node in the convolution layer extracts features from the input images by performing a convolution operation on the input nodes. The max-pooling layer abstracts the features by averaging or calculating the maximum value of input nodes [25,26]. The DenseNet169 is a highly supervised network with a 5-layer dense block, the standard ResNet structure, and a growth rate of $k = 4$. A dense DenseNet block is suitable for object detection because the output of each layer includes the output of all layers before it and includes both low-level and high-level image features [27].

The DenseNet169 was trained on the ILSVRC 2012 classification dataset containing 1.2 million images and 1,000 classes. Before being used as input for the DenseNet169, the images from the dataset were cropped to the size of the DenseNet169, introducing a novel connectivity pattern that included direct connections between any layer and the layers below it to enhance information flow further [28]. The first layer of the DenseNet169 takes all feature maps from the layers before it as input, as shown in Fig. 3, by where a single tensor is and the layers' concatenated features. Convolution and average pooling were used as transition layers between adjacent dense blocks in the DenseNet169. A global average pooling was done at the end of the last dense block, and a Soft-

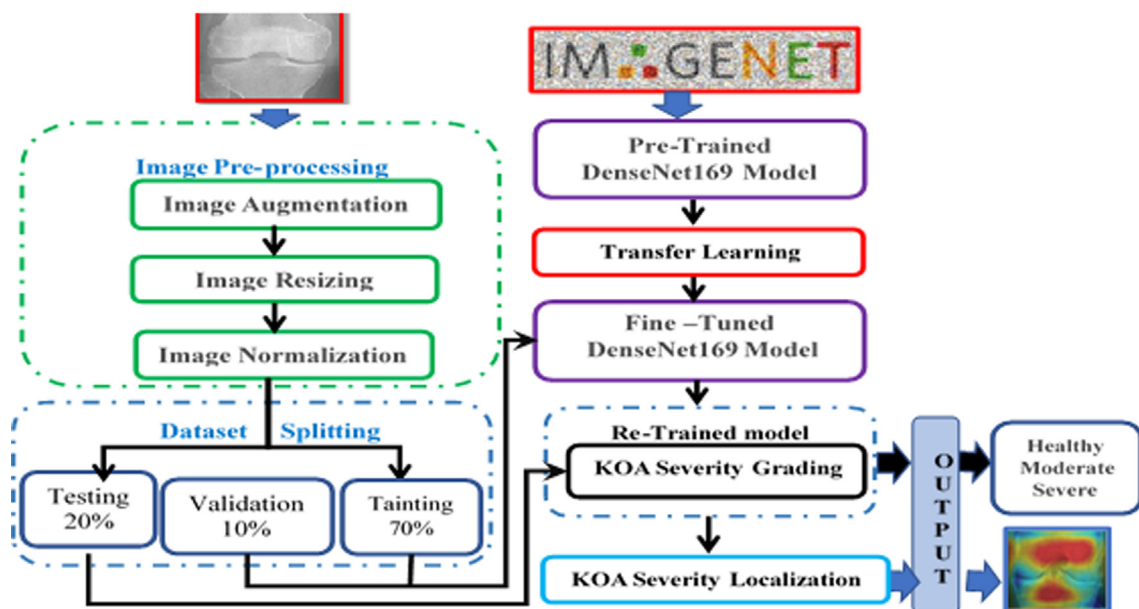


Fig. 2. The overall model architecture.

max classifier was connected. Fine-tuning was applied to the DenseNet169 to unfreeze the different layers of the DenseNet169 and make it a trainable model. The structure of DenseNet-169 in our proposed model is shown in Fig. 3, in Table 2 and in Table 3.

Where patience specifies how many epochs without improvement before the learning rate is adjusted, stop_patience specifies how many times to adjust the learning rate without improvement to stop training, and factor is used to reduce the learning rate.

3.2.3. KOA localization and progression analysis using Grad-CAM

Rationalizing and explaining the results in various DL models related to medical imaging is essential. Selvaraju et al. [29] showed a Grad-CAM method that makes DL models easy to understand and could be used to create a visual explanation for a DL technique to find out more about the DL model while it is being classified. The Grad-CAM algorithm was used to superimpose heat maps of X-ray dataset images on our proposed model, as depicted in Fig. 4. The Grad-CAM algorithm was used to examine images of three classes: healthy, moderate, and severe classes. There is no opacity, which differentiates other patients from normal patients on normal X-rays. No significant region was localized in normal X-rays, as shown in Fig. 4. Our model can identify the localized regions with bilateral multifocal ground-glass opacities (GGO) in KOA disease by looking at the generated heatmaps.

4. Model implementation and evaluation

4.1. Model evaluation metrics

Eqs. (3) to (7), namely, accuracy, precision, sensitivity, specificity, and F1 score, were used to evaluate the proposed model.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (3)$$

$$\text{Precision} = TP / (TP + FP) \quad (4)$$

$$\text{Sensitivity} = TP / (TP + FN) \quad (5)$$

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

$$\text{F1 - score} = 2 * (Precision * Sensitivity) / (Precision + Sensitivity) \quad (7)$$

TP means true positives, TN means true negatives, FP means false positives, and FN means false negatives. The F1-score is the weighted average of recall (sensitivity) and precision. TP will give the count of the quantity of positive examples that were accurately delegated positives. TN will give the count of the quantity of neg-

Table 2

The DenseNet 169 model architectures.

Layers	Output Size	DenseNet 169
Conv	112 × 112	stride 2, 7 × 7 conv
Pooling	56 × 56	stride 2, 3 × 3 max pool
Dense Block (1)	56 × 56	[1 × 1 conv, 3 × 3 conv] × 6
Transition (1) Layer	56 × 56	1 × 1 conv
	28 × 28	stride 2, 2 × 2 average pool
Dense Block (2)	28 × 28	[1 × 1 conv, 3 × 3 conv] × 12
Transition (2) Layer	28 × 28	1 × 1 conv
	14 × 14	stride 2, 2 × 2 average pool
Dense Block (3)	14 × 14	[1 × 1 conv, 3 × 3 conv] × 32
Transition (3) Layer	14 × 14	1 × 1 conv
	7 × 7	stride 2, 2 × 2 average pool
Dense Block (4)	7 × 7	[1 × 1 conv, 3 × 3 conv] × 32
Classifier (Classification Layer)	1 × 1	7 × 7 global average pool
	1000	Softmax, 1000D fully-connected

Table 3

The network architecture.

parameter	Value
epochs	20
Learning rate	0.001
Batch size	32
patience	1
stop_patience	10
factor	0.5
padding	zero_padding

ative examples that were accurately delegated negatives. FP is the count of number of negative samples that were classified as positive samples. FN is the count of number of positive samples that were classified as negative samples.

4.2. Model implementation

In the implementation process, the OAI dataset was split into 70% for the training, 20% for the testing, and 10% for the validation process. In the Kaggle environment, the implementation process was carried out. The characteristics of the PC used for the implementation process are listed in Table 4:

Table 5 shows the results from the DenesNet169 and five other DL models: InceptionV3, Xception, ResNet50, DenseNet121, and InceptionResnet, applied to the OAI dataset for the multi-classification task. Additionally, the highest values that were measured are highlighted in Bold. Table 5 shows that DenseNet169, InceptionV3, Xception, ResNet50, DenseNet121, and InceptionRes-

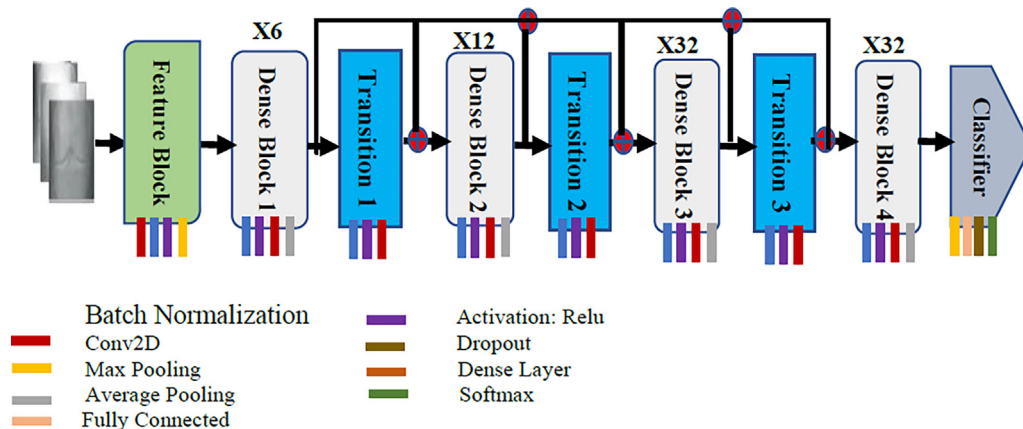


Fig. 3. Three types of the DenseNet-169 modules from left to the right.

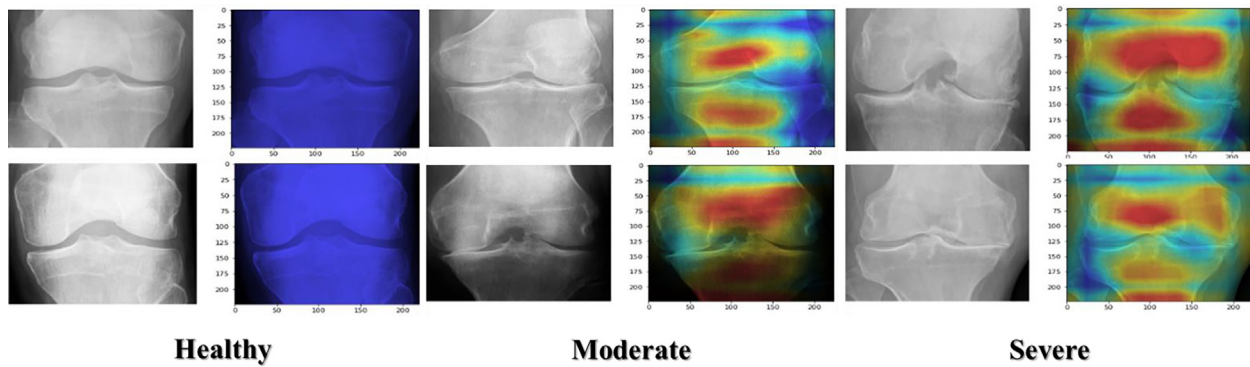


Fig. 4. Grad-CAM images of parts of the test knee in healthy, moderate, and severe classes.

netV2 offered an accuracy 95.93%, 94.97%, 94.32%, 93.64%, 95.21%, 93.40%, respectively. The DenseNet169 model had the highest accuracy, sensitivity, specificity, precision, and F1-score, 95.93%, 88.77%, 95.41%, 85.8%, and 87.08%, respectively. ResNet50 and InceptionResNetV2 had the lowest accuracy, 93.64%, and 93.40%, respectively. InceptionV3 and Xception had the lowest sensitivity, 85.56%, and 86.17%, respectively. InceptionV3 and Xception had the lowest specificity, 92.58% and 81.01%, respectively. Xception

and ResNet50 had the lowest precision, 81.01% and 80.41%, respectively, and Xception and ResNet50 had the lowest F1-score, 83.18%, and 83.16%, respectively.

Table 6 shows the results of the multi-classification process in detail. In this process, we classified three classes: healthy, moderate, and severe. The results obtained from implementing the six CNN models: DenseNet169, InceptionV3, Xception, ResNet50, DenseNet121, and InceptionResnet, on the OAI dataset for measuring accuracy, precision, sensitivity, specificity, and F1-score. The highest values are highlighted in red for the severe class, green for the moderate class, and blue for the healthy class.

For the severe class, precision, recall, F1-Score, accuracy, and specificity are 86%, 84.31%, 85.15%, 99.09%, and 99.56% for DenseNet169, respectively. DenseNet169, DenseNet121, InceptionV3, and InceptionResNetV2 achieved the highest accuracy 99.09%, 99.28%, 99.21%, and 99.15%, respectively. DenseNet169, DenseNet121, InceptionV3, and ResNet50 achieved the highest sensitiv-

Table 4

The used PC specifications.

Central Processing Unit	Intel(R) Core (TM) i7, 2.30 GHz
RAM	16 GBs
Graphical Processing Unit Type	16 GB for NVIDIA
OS	Windows, x32

Table 5

The results of the multi-classification for the OAI dataset.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)
DenseNet169	95.93	88.77	95.41	85.80	87.08
InceptionV3	94.97	85.56	92.58	85.37	85.38
Xception	94.32	86.17	81.01	81.01	83.18
ResNet50	93.64	87.77	94.58	80.41	83.16
DenseNet121	95.21	88.71	94.59	85.49	86.77
InceptionResNetV2	93.40	86.49	94.03	82.73	83.46

Table 6

The multi-classification results for three classes.

Model	Metric	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	Specificity (%)
DenseNet169	Healthy	98.29	95.44	96.84	94.81	91.61
	Moderate	73.11	86.55	79.26	93.90	95.05
	Severe	86.00	84.31	85.15	99.09	99.56
InceptionV3	Healthy	96.62	95.22	95.92	93.24	83.21
	Moderate	69.92	77.13	73.35	92.45	94.84
	Severe	89.58	84.31	86.87	99.21	99.69
Xception	Healthy	97.80	93.20	95.44	92.57	89.42
	Moderate	64.46	82.96	72.55	91.55	92.88
	Severe	80.77	82.35	81.55	98.85	99.38
ResNet50	Healthy	98.59	91.10	94.70	91.49	93.43
	Moderate	59.94	87.89	71.27	90.46	90.86
	Severe	82.69	84.31	83.50	98.97	99.44
DenseNet121	Healthy	97.97	94.21	96.05	93.54	90.15
	Moderate	68.71	85.65	76.25	92.81	93.93
	Severe	89.80	86.27	88.00	99.28	99.69
InceptionResNetV2	Healthy	98.28	90.74	94.36	90.94	91.97
	Moderate	58.81	88.34	70.61	90.10	90.37
	Severe	91.11	80.39	85.42	99.15	99.75

ity 84.31%, 86.27%, 84.31%, and 84.31% respectively. InceptionResNetV2 and DenseNet121 achieved the highest specificity, 99.75%, 99.69%, and precision, 91.11% and 89.8%, respectively. DenseNet121 and InceptionV3 achieved the highest F1-score, 86.87%, and 88%, respectively.

For the moderate class, precision, recall, F1-Score, accuracy, and specificity are 73.11%, 86.55%, 79.26%, 93.90%, and 95.05% for DenseNet169, respectively. DenseNet169, InceptionV3, DenseNet121 achieved the highest accuracy 93.90%, 92.45%, and 92.81% respectively. InceptionResNetV2 and ResNet50 achieved the highest sensitivity, 88.34% and 87.89%, respectively. DenseNet169 and DenseNet121 achieved the highest specificity, 95.05%, and 93.93%, respectively. DenseNet169 achieved the highest precision, 73.11%. DenseNet169 and DenseNet121 achieved the highest F1-score, 79.26%, and 76.25%, respectively.

For the healthy class, precision, recall, F1-Score, accuracy, and specificity are 98.29%, 95.44%, 96.84%, 94.81%, and 91.61% for DenseNet169, respectively. DenseNet169 and DenseNet121 achieved the highest accuracy, 94.81%, and 93.54%, respectively. DenseNet169 and InceptionV3 achieved the highest sensitivity, 95.44%, and 95.22%, respectively. DenseNet169, ResNet50, and InceptionResNetV2 achieved the highest specificity 91.61%, 93.43%, and 91.97%, respectively. DenseNet169, ResNet50, and InceptionResNetV2 achieved the highest precision 98.29%, 98.59%, and 98.28%, respectively. DenseNet169 and DenseNet121 achieved the highest F1-score, 96.84%, and 96.05%, respectively.

The loss and accuracy for each epoch of the validation and the training datasets are depicted in Fig. 5. At epoch 20, the DenseNet169 model's training and validation accuracy were over 90%, and the DenseNet169 model's training and validation loss were similar to each other at 1. The training accuracy of the InceptionV3 model was more than 95%, and the validation accuracy ranged between 90% and 95%. The InceptionV3 model's training and validation loss were more than 2. The training accuracy of the Xception model was more than 90%, and the validation accuracy ranged between 80% and 90%. The Xception model's training loss was less than 2, and the validation loss was more than 2. The training accuracy of the ResNet50 model ranged between 90 and 95%, and the validation accuracy was less than 90%. The ResNet50 model's training and validation loss were less than 5. The training and validation accuracy of the DenseNet121 ranged between 90 and 95%, and the training and validation loss ranged between 1 and 2. The training accuracy of the InceptionResNetV2 model was more than 90%, and the validation accuracy of the InceptionResNetV2 model ranged between 85 and 90%. The training loss of the InceptionResNetV2 was less than 4, and the validation loss of the InceptionResNetV2 ranged between 4 and 5.

Fig. 6 shows the confusion matrix for the six CNN models in the Healthy, Moderate, and Severe Classes test dataset. The test dataset was classified into three classes: Healthy class with 1382 images, Moderate class with 223 images, and Severe class with 51 images. For class healthy, the accuracy of the DenseNet169 and the Incep-

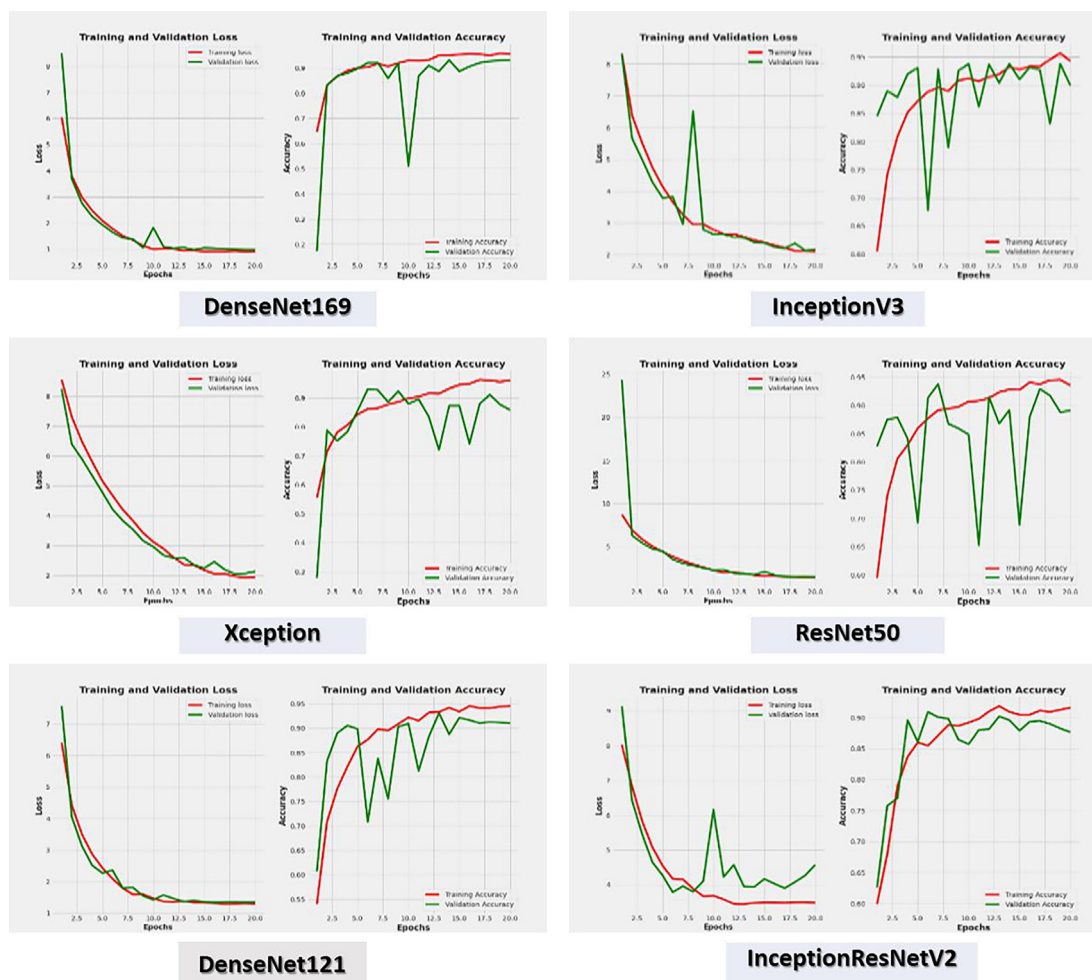


Fig. 5. Each epoch's six CNN models' loss and accuracy.

tionV3 models was 95%, as they predicted 1319 images correctly from 1382. The accuracy of the Xception model was 93%, as it predicted 1288 images correctly. The accuracy of the ResNet50 model was 91%, as it predicted 1259 images correctly. The accuracy of the DenseNet121 model was 94% as it predicted 1302 images correctly, and the accuracy of the InceptionResNetV2 model was 90% as it predicted 1254 images correctly.

For class moderate, the accuracy of the DenseNet169 model was 86.5%, as it predicted 193 images correctly from 223. The accuracy of the InceptionV3 model was 77%, as it predicted 172 images correctly. The accuracy of the Xception model was 82%, as it predicted 185 images correctly. The accuracy of the ResNet50 model was 87% as it predicted 196 images correctly, the accuracy of the DenseNet121 model was 85% as it predicted 191 images correctly, and the accuracy of the InceptionResNetV2 model was 88% as it predicted 197 images correctly.

For class severe, the accuracy of DenseNet169, InceptionV3, and ResNet50 models was 84%, as they predicted 43 images correctly from 51. The accuracy of the Xception model was 82%, as it predicted 42 images correctly. The accuracy of the DenseNet121 model was 86%, as it predicted 44 images correctly. The accuracy of the InceptionResNetV2 model was 80%, as it predicted 41 images correctly.

After we experimented with multi-classification, we experimented with binary classification. The binary classification results obtained from the six CNN models performed on the OAI dataset for the binary classification are presented in Table 7. The highest values that were measured are highlighted in Bold. DenseNet169, InceptionV3, Xception, ResNet50, DenseNet121, and InceptionResNetV2 had an accuracy of 93.78%, 92.39%, 93.07%, 89.79%, 91.68%, and 90.70% respectively, as shown in Table 7. The accuracy, sensitivity, specificity, precision, and F1-score of the DenseNet169 model were the highest, with 93.78%, 91.29%, 91.29%, 87.57%, and 89.27%, respectively. With 89.79%, ResNet50 had the lowest accuracy. The precision of ResNet50 and InceptionResNetV2 was the lowest, at 81.93% and 81.89%, respectively. The sensitivity and specificity of ResNet50 were the lowest, at 80.133%. The F1-score for ResNet50 was the lowest, at 80.99%.

Table 8 provides a comprehensive breakdown of the binary classification process's results. We divided the OAI dataset into two classes: positive and negative KOA. The precision, sensitivity, f1-score, accuracy, and specificity were measured using the six CNN models: DenseNet169, InceptionV3, Xception, ResNet50, and DenseNet121 on the OAI dataset. The red color is used to highlight the highest values for the Negative KOA class, and the green color is used to highlight the highest values for the positive KOA class.

DenseNet169				InceptionV3				Xception						
True Label	Healthy	1319	63	0	True Label	Healthy	1316	66	0	True Label	Healthy	1288	94	0
	Moderate	23	193	7		Moderate	46	172	5		Moderate	28	185	10
	Severe	0	8	43			Severe	0	8			43	Severe	1
		Healthy	Moderate	Severe				Healthy	Moderate	Severe				Healthy
		Predicted Label					Predicted Label					Predicted Label		

ResNet50				DenseNet121				InceptionResNetV2						
True Label	Healthy	1259	123	0	True Label	Healthy	1302	80	0	True Label	Healthy	1254	128	0
	Moderate	18	196	9		Moderate	27	191	5		Moderate	22	197	4
	Severe	0	8	43			Severe	0	7			44	Severe	0
		Healthy	Moderate	Severe				Healthy	Moderate	Severe				Healthy
		Predicted Label					Predicted Label					Predicted Label		

Fig. 6. Test dataset's confusion matrix for the six CNN models for Healthy, Moderate, and Severe Classes.

Table 7

The results of the binary classification for the OAI dataset.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)
DenseNet169	93.78	91.29	91.29	87.57	89.27
InceptionV3	92.39	86.07	86.07	86.29	86.18
Xception	93.07	86.21	86.21	88.92	87.52
ResNet50	89.79	80.13	80.13	81.93	80.99
DenseNet121	91.68	88.71	88.71	93.95	85.20
InceptionResNetV2	90.70	90.62	89.10	81.89	85.19

For the negative KOA class, precision, sensitivity, F1-score, accuracy, and specificity are 97.48%, 95.01%, 96.23%, 93.78%, and 87.59% for DenseNet169, respectively. DenseNet169 and DenseNet121 had the highest accuracy rates of 93.78% and 94.50%, respectively. The precision of DenseNet169 and InceptionResNetV2 was the highest, with 97.48% and 97.70%, respectively. The highest sensitivity was achieved by InceptionV3 and DenseNet121, with 95.51% and 97.25%, respectively. DenseNet169 and Xception had the highest specificity, with 87.59% and 84.67%, respectively. The F1-score of DenseNet169 and DenseNet121 were the highest, with 96.23% and 96.83%, respectively.

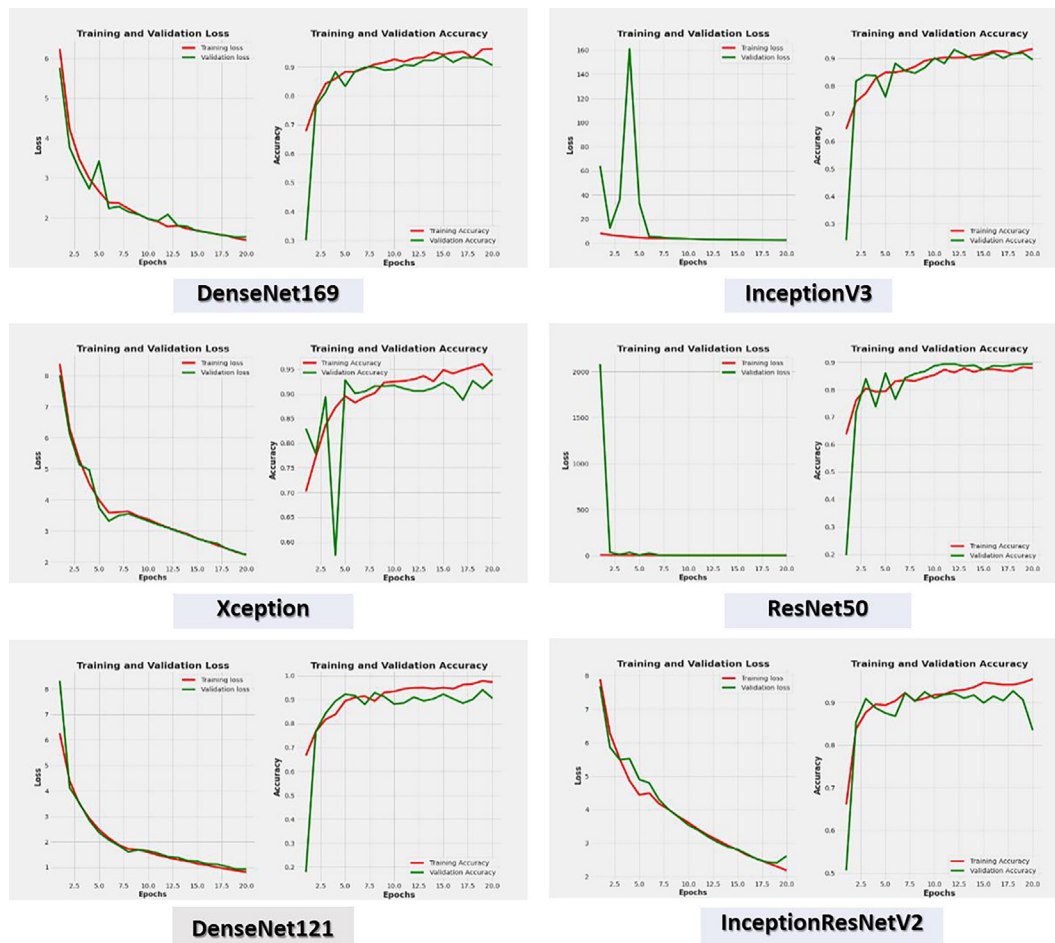
Table 8

The binary classification results for two classes.

Model	Class	Precision (%)	Sensitivity (%)	F1-score (%)	Accuracy (%)	Specificity (%)
DenseNet169	Negative KOA	97.48	95.01	96.23	93.78	87.59
	Positive KOA	77.67	87.59	82.33	93.78	95.00
InceptionV3	Negative KOA	95.38	95.51	95.44	92.39	76.64
	Positive KOA	77.21	76.64	76.92	92.39	95.51
Xception	Negative KOA	96.79	91.75	94.21	90.57	84.67
	Positive KOA	67.05	84.67	74.84	90.57	91.75
ResNet50	Negative KOA	93.29	94.57	93.93	89.79	83.94
	Positive KOA	70.59	65.69	68.05	89.79	95.07
DenseNet121	Negative KOA	96.41	97.25	96.83	94.50	81.75
	Positive KOA	85.50	81.75	83.58	94.50	97.25
InceptionResNetV2	Negative KOA	97.70	90.73	96.11	90.53	76.48
	Positive KOA	65.37	90.54	80.86	90.53	94.73

For the positive KOA class, precision, sensitivity, F1-score, accuracy, and specificity are 77.67%, 87.59%, 82.33%, 93.78%, and 95% for DenseNet169, respectively. DenseNet169 and DenseNet121 had the highest accuracy rates of 93.78% and 94.50%, respectively. DenseNet121 had the highest precision, sensitivity, specificity, and F1-score rates of 85.50%, 90.54%, 97.25%, and 83.58%, respectively.

Fig. 7 depicts the accuracy and loss of the training and validation data for each epoch. At epoch 20, the DenseNet169 model's training and validation accuracy were over 90%, and the DenseNet169 model's training and validation loss were less than 2. The training accuracy of the InceptionV3 model was more than 90%, and the validation accuracy was 90%. The InceptionV3 model's

**Fig. 7.** The six CNN models' loss and accuracy for each epoch.

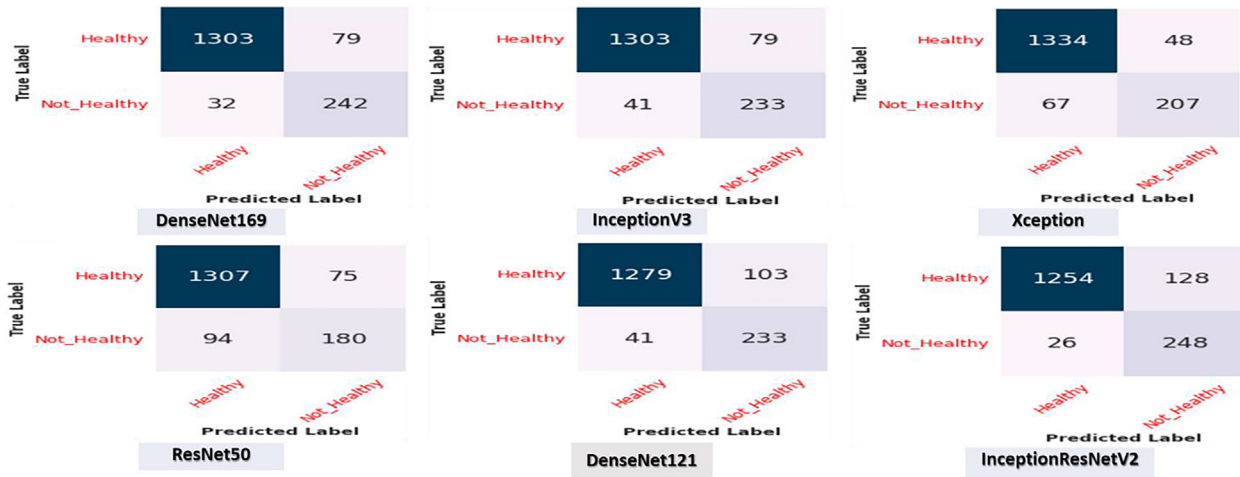


Fig. 8. Test dataset's confusion matrix for the six CNN models for positive (not healthy) and negative (healthy) KOA classes.

Table 9

The comparison of the proposed model with the state-of-art methods.

Reference	Methodology	performance
Schiratti et al. [12]	Deep learning	AUC 65%
Wang et al. [13]	CNN and YOLO	Accuracy: 69.18%.
Kondal et al. [14]	CNN	Accuracy:87%
Tiulpin et al. [15]	CNN	Accuracy: 66.71% and ROC curve of 0.93
Kokkotis et al. [16]	Random Forest	Accuracy: 73.55%
McCabe et al. [17]	LR and Cox regression	AUC: 0.67 for the validation data and 0.75 for the test data.
Liu et al. [18]	RPN and Fast R-CNN	sensitivity above 78% and specificity above 94%
Lim et al. [19]	CNN	AUC: 76.8%
Tiulpin et al. [20]	CNN	Precision: 0.62 and AUC: 0.8
Tolpadi et al. [21]	deep siamese CNN and a refined ResNet-34	Accuracy: 61%
The Proposed Model	DenseNet169	Accuracy: 95.93%

training and validation loss were close to 0. The training accuracy of the Xception model was more than 95%, and the validation accuracy ranged between 90% and 95%. The Xception model's training and validation loss ranged between 2 and 3. The training and validation accuracy of the ResNet50 model was close to 90%. The ResNet50 model's training and validation loss were 0. The training and validation accuracy of the DenseNet121 ranged between 90 and 100%. The training loss of the DenseNet121 was less than 1, and the validation loss was 1. The training accuracy of the InceptionResNetV2 model was more than 90%, and the validation accuracy of the InceptionResNetV2 model ranged between 80 and 90%. The training and validation loss of the InceptionResNetV2 ranged between 2 and 3.

Fig. 8 shows the confusion matrix for the six CNN models in the test dataset. The test dataset was classified into two classes: The negative (healthy) class with 1382 images and the Positive (not healthy) class with 274 images. DenseNet169, InceptionV3, and ResNet50 correctly predicted 1303 out of 1382 from the negative class, while the Xception model was 96.5% accurate in predicting 1334 negative images, respectively. DenseNet121 was 92.5% accurate in predicting 1279 negative images, and InceptionResNetV2 was 90.7% accurate in predicting 1254 negative images. DenseNet169's accuracy for the Positive class was 88%, correctly predicting 242 images out of 274, while the accuracy of InceptionV3 and DenseNet121 was 85%, correctly predicting 233 images. The Xception model correctly predicted 207 images with an accuracy of 75.5%, the ResNet50 model correctly predicted 180 images with an accuracy of 65.6%, and the InceptionResNetV2 model accurately predicted 248 images with an accuracy of 90.5%.

4.3. Model's result comparison with the literature

Regarding accuracy for the multi-classification process, the results demonstrated that our proposed model with DenseNet169 performed better than the most recent methods listed in Table 9. It achieved a remarkable accuracy of 95.93%. Additionally, from the patient's perspective, our model would improve the procedure by lowering the cost of diagnosis, speeding up diagnosis, and delaying disease progression. The proposed model would successfully localize the opacities' peripheral, diffuse, and vascular thickening distributions and be used to determine the severity of KOA diseases through multi-classification of KOA severity. Consequently, this localization would enable clinicians to comprehensively comprehend the primary causes of KOA using the proposed model. Compared to other ongoing frameworks, the evaluation of the proposed model revealed that it is an unparalleled perceptive outcome with tuning.

5. Conclusion

This research proposed a fine-tuning KOA diagnosis model using DL techniques. The model was proposed to enhance the efficiency of KOA diagnosis. The proposed model will reduce the cost of diagnosis, speed up diagnosis, and delay disease progression, enhancing the procedure from the patient's perspective. The proposed model will determine the degree of KOA diseases by making multi-classification and binary classifications of the KOA severity. It will successfully localize the opacities' peripheral, diffuse distri-

bution, and vascular thickening. Therefore, the proposed model may enable clinicians to understand the primary causes of KOA through this localization. The proposed model was evaluated over the OAI dataset. The OAI dataset was pre-processed by artifact removal, resizing, contrast handling, and a normalization technique. The DenseNet169 was applied over the OAI dataset and compared with five DL algorithms: InceptionV3, Xception, ResNet50, DenseNet121, and InceptionResNetV2. In multi-classification, the DenseNet169 model achieved 95.93%, 88.77%, 95.41%, 85.8%, and 87.08%, for accuracy, sensitivity, specificity, precision, and F1-score, respectively, and in binary classification, the DenseNet169 model achieved 93.78%, 91.29%, 91.29%, 87.57%, and 89.27%, for accuracy, sensitivity, specificity, precision, and F1-score, respectively. The evaluation of the proposed model showed that it is an unrivaled perceptive outcome with tuning as opposed to other ongoing existing frameworks. However, the training and testing time phases are the drawbacks of our proposed model. This drawback was due to the structure of the model. Therefore, in future work, we will simplify our model to reduce the time of training and testing phases.

CRedit authorship contribution statement

Sameh Abd El-Ghany: Conceptualization, Methodology, Writing – review & editing, Supervision. **Mohammed Elmogy:** Conceptualization, Methodology, Writing – review & editing. **A.A. Abd El-Aziz:** Data curation, Writing – original draft, Writing – review & editing.

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.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at Jouf University for funding this work through research grant No (DSR2020-04-3649).

References

- [1] J. Antony, K. McGuinness, K. Moran, and N. E. O'Connor, "Feature learning to automatically assess radiographic knee osteoarthritis severity," *CoRR, ArXiv*, vol. abs/1908.08840, 2019.
- [2] Woolf D. The bone and joint decade. Strategies to reduce the burden of disease: the Bone and Joint Monitor Project. *J Rheumatol* 2003;67:6–9.
- [3] Deshpande BR, Katz JN, Solomon DH, Yelin EH, Hunter DJ, et al. Number of persons with symptomatic knee osteoarthritis in the US: impact of race and ethnicity, age, sex, and obesity. *Arthritis Care Res (Hoboken)* 2016;68(12):1743–50.
- [4] Vina ER, Kwok CK. Epidemiology of osteoarthritis: literature update. *Curr Opin Rheumatol* 2018;30(2):160–7.
- [5] J. Hermans J, M. A. Koopmanschap, S. M. A. Bierma-Zeinstra, J. H. van Linge, J. A. N. Verhaar et al., Productivity costs and medical costs among working patients with knee osteoarthritis, *Arthritis Care Res (Hoboken)*, vol. 64, no. 3, pp: 853–861, 2012.
- [6] Maiese K. Picking a bone with wisp1 (ccn4): new strategies against degenerative joint disease. *J Transl Sci* 2016;1(3):83–5.
- [7] Losina E, Paltiel AD, Weinstein AM, Yelin E, Hunter DJ, et al. Lifetime medical costs of knee osteoarthritis management in the United States: impact of extending indications for total knee arthroplasty. *Arthritis Care Res* 2015;67(2):203–15.
- [8] Kellgren J, Lawrence J. Radiological assessment of osteo-arthritis. *Ann Rheum Dis* 1957;16(4):494–502.
- [9] Kohn MD, Sassoon AA, Fernando ND. Classifications in brief: kellgren-lawrence classification of osteoarthritis. *Clin Orthop Relat Res* 2016;474(8):1886–93.
- [10] Wenham CYJ, Grainger AJ, Conaghan PG. The role of imaging modalities in the diagnosis, differential diagnosis and clinical assessment of peripheral joint osteoarthritis. *Osteoarthritis Cartil* 2014;22(10):1692–702.
- [11] Braun HJ, Gold GE. Diagnosis of osteoarthritis: imaging. *Bone* 2012;51(2):278–88.
- [12] Schiratti JB, Dubois R, Herent P, Cahané D, Dachary J, et al. A deep learning method for predicting knee osteoarthritis radiographic progression from MRI. *Arthritis Res Ther* 2021;23(1):262.
- [13] 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. 5586529, 2021.
- [14] Kondal S, Kulkarni V, Gaikwad A, Kharat A, Pant A, et al. Automatic grading of knee osteoarthritis on the kellgren-lawrence scale from radiographs using convolutional neural networks. *Advances in Deep Learning, Artificial Intelligence and Robotics Lecture Notes in Networks and Systems* 2022;249:163–73.
- [15] Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S, et al. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. *Sci Rep* 2018;8(1):1727.
- [16] Kokkoti C, Ntakolia C, Moustakidis S, Giakas G, Tsaopoulos D, et al. Explainable machine learning for knee osteoarthritis diagnosis based on a novel fuzzy feature selection methodology. *Phys Eng Sci Med* 2022;45(1):219–29.
- [17] McCabe PG, Lisboa P, Baltzopoulos B, Olier I. Externally validated models for first diagnosis and risk of progression of knee osteoarthritis. *PLoS One* 2022;17(7):e0270652.
- [18] Liu B, Luo J, Huang H. Toward automatic quantification of knee osteoarthritis severity using improved faster R-CNN. *Int J Comput Assist Radiol Surg* 2020;15(3):457–66.
- [19] Lim J, Kim J, Cheon S. A deep neural network-based method for early detection of osteoarthritis using statistical data. *Int J Environ Res Public Health* 2019;16(7):1281.
- [20] Tiulpin A, Klein S, Bierma-Zeinstra SMA, Jérôme T, Esa R, et al. Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. *Sci Rep* 2019;9(1):20038.
- [21] Tolpadi AA, Lee J, Padoia V, Majumdar S. Deep learning predicts total knee replacement from magnetic resonance images. *Sci Rep* 2020;10(1):6371.
- [22] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis* 2015;115:211–52.
- [23] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2009;22:1345–59.
- [24] OpenCV: "Image thresholding". [Online]. Available: https://docs.opencv.org/master/d7/d4d/tutorial_py_thresholding.html (Last access on 26/11/2022).
- [25] Adnan M, Rahman F, Imrul M, Zahid NAL, Shabnam S. Handwritten bangla character recognition using inception convolutional neural network. *Int J Comput Appl* 2018;181(17):48–59.
- [26] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86(11):2278–324.
- [27] P. Zhou, B. Ni, C. Geng, J. Hu, and Y. Xu, "Scale-transferrable object detection," in *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 528–537, 2018.
- [28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 2261–2269, 2017.
- [29] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, et al., "Grad-cam: visual explanations from deep networks via gradient-based localization," in *Proc. 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 618–626, 2017.

Sameh Abd El-Ghany received B.S. degree in (2003) from information systems department, Mansoura University, Egypt, and the M.Sc. degree in (2009) from Mansoura University, Egypt, entitled thesis 'an approach for image retrieval based on mobile agent. He received Ph.D. degree from faculty of computer science and information, Mansoura University, Egypt. His research interests include information retrieval, image processing, sentiment analysis and semantic web. He is currently an Assistant Professor at the Department of information systems, college of information and computer sciences, jouf University, Saudi Arabia.

A. A. Abd El-Aziz have completed Ph.D. degree in June 2014 in Information Science Technology from Anna University, Chennai-25, India. He has received B.Sc., and Master of Computer Science in 1999 and 2006 respectively from Faculty of Science, Cairo University. Now, he is an Ass. Prof in the FGSSR, Cairo University, Egypt. He has 21 years' experience in teaching at Cairo University, Egypt. His research interests are database system, database security, Cloud computing, Big data, ML and XML security. He has authored/coauthored over 68 research publications in peer-reviewed reputed journals and conference proceedings. He advised more than 10 master's graduates. Moreover, he has also served as a technical program committee member for many workshops and conferences and he has served as a reviewer for various international journals.

Mohamed Elmogy (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the Faculty of Engineering, Mansoura University, Mansoura, Egypt, and the Ph.D. degree from the Department of Informatics, MIN Faculty, University of Hamburg, Hamburg, Germany, in 2010. He worked as a Visiting Researcher with the Department of Bioengineering, University of Louisville, Louisville, KY, USA, from July 2016 to August 2019. He is currently a Professor and the Chair of the Department of Information Technology, Faculty of Computers and Information, Mansoura University. He has authored/coauthored over 220 research publications in peer-

reviewed reputed journals, book chapters, and conference proceedings. He advised more than 30 master's and Ph.D. graduates. His current research interests include computer vision, medical image analysis, machine learning, pattern recognition,

and biomedical engineering. He is also a Professional Member of the ACM Society. He has also served as a technical program committee member for many workshops and conferences. He has served as a reviewer for various international journals.