

World Journal of *Orthopedics*

World J Orthop 2022 June 18; 13(6): 544-621



MINIREVIEWS

- 544 Management of hip fracture in COVID-19 infected patients
Qin HC, He Z, Luo ZW, Zhu YL
- 555 Topical use of tranexamic acid: Are there concerns for cytotoxicity?
Gkias I, Kontokostopoulos AP, Tsirigkakis SE, Kostas-Agnantis I, Gelalis I, Korompilias A, Pakos E
- 564 Managements of osteoporotic vertebral compression fractures: A narrative review
Patel D, Liu J, Ebraheim NA
- 574 Hip prosthetic loosening and periprosthetic osteolysis: A commentary
Mjöberg B

ORIGINAL ARTICLE**Basic Study**

- 578 Reducing bacterial adhesion to titanium surfaces using low intensity alternating electrical pulses
Bernaus M, Guillem-Martí J, Bermúdez-Castel A, Calero JA, Torres D, Veloso M, Font-Vizcarra L
- 587 Content of blood cell components, inflammatory cytokines and growth factors in autologous platelet-rich plasma obtained by various methods
Dejneke M, Witkowski J, Moreira H, Placzowska S, Morasiewicz P, Reichert P, Królikowska A

Retrospective Study

- 603 Evaluation of artificial intelligence models for osteoarthritis of the knee using deep learning algorithms for orthopedic radiographs
Tiwari A, Poduval M, Bagaria V
- 615 Role of joint aspiration before re-implantation in patients with a cement spacer in place
Huguet S, Bernaus M, Gómez L, Cuchi E, Soriano A, Font-Vizcarra L

ABOUT COVER

Editorial Board Member of *World Journal of Orthopedics*, Alaaeldin A Ahmad, MBChB, MD, Adjunct Professor, Surgeon, Department of Orthopedic, Palestine Polytechnic University, Ramallah 3985, West Bank, Palestine. alaaahmad@hotmail.com

AIMS AND SCOPE

The primary aim of *World Journal of Orthopedics* (*WJO*, *World J Orthop*) is to provide scholars and readers from various fields of orthopedics with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

WJO mainly publishes articles reporting research results and findings obtained in the field of orthopedics and covering a wide range of topics including arthroscopy, bone trauma, bone tumors, hand and foot surgery, joint surgery, orthopedic trauma, osteoarthropathy, osteoporosis, pediatric orthopedics, spinal diseases, spine surgery, and sports medicine.

INDEXING/ABSTRACTING

The *WJO* is now abstracted and indexed in PubMed, PubMed Central, Emerging Sources Citation Index (Web of Science), Scopus, Reference Citation Analysis, China National Knowledge Infrastructure, China Science and Technology Journal Database, and Superstar Journals Database. The 2021 edition of Journal Citation Reports® cites the 2020 Journal Citation Indicator (JCI) for *WJO* as 0.66. The *WJO*'s CiteScore for 2020 is 3.2 and Scopus CiteScore rank 2020: Orthopedics and Sports Medicine is 87/262.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: Ying-Yi Yuan, Production Department Director: Xiang Li, Editorial Office Director: Jin-Lei Wang.

NAME OF JOURNAL

World Journal of Orthopedics

ISSN

ISSN 2218-5836 (online)

LAUNCH DATE

November 18, 2010

FREQUENCY

Monthly

EDITORS-IN-CHIEF

Massimiliano Leigheb

EDITORIAL BOARD MEMBERS

<http://www.wjgnet.com/2218-5836/editorialboard.htm>

PUBLICATION DATE

June 18, 2022

COPYRIGHT

© 2022 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS

<https://www.wjgnet.com/bpg/gerinfo/204>

GUIDELINES FOR ETHICS DOCUMENTS

<https://www.wjgnet.com/bpg/GerInfo/287>

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH

<https://www.wjgnet.com/bpg/gerinfo/240>

PUBLICATION ETHICS

<https://www.wjgnet.com/bpg/GerInfo/288>

PUBLICATION MISCONDUCT

<https://www.wjgnet.com/bpg/gerinfo/208>

ARTICLE PROCESSING CHARGE

<https://www.wjgnet.com/bpg/gerinfo/242>

STEPS FOR SUBMITTING MANUSCRIPTS

<https://www.wjgnet.com/bpg/GerInfo/239>

ONLINE SUBMISSION

<https://www.f6publishing.com>



Retrospective Study

Evaluation of artificial intelligence models for osteoarthritis of the knee using deep learning algorithms for orthopedic radiographs

Anjali Tiwari, Murali Poduval, Vaibhav Bagaria

Specialty type: Orthopedics

Provenance and peer review:

Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0

Grade B (Very good): B

Grade C (Good): C

Grade D (Fair): 0

Grade E (Poor): 0

P-Reviewer: Labusca L, Romania;
Tanabe S, Japan

Received: November 30, 2021

Peer-review started: November 30, 2021

First decision: January 11, 2022

Revised: January 20, 2022

Accepted: May 13, 2022

Article in press: May 13, 2022

Published online: June 18, 2022



Anjali Tiwari, Vaibhav Bagaria, Department of Orthopedics, Sir H. N. Reliance Foundation Hospital and Research Centre, Mumbai 400004, India

Murali Poduval, Lifesciences Engineering, Tata Consultancy Services, Mumbai 400096, India

Vaibhav Bagaria, Department of Orthopedics, Columbia Asia Hospital, Mumbai 400004, India

Corresponding author: Vaibhav Bagaria, FCPS, MBBS, MS, Director, Department of Orthopedics, Sir H. N. Reliance Foundation Hospital and Research Centre, Raja Rammohan Roy Road, Prarthana Samaj, Mumbai 400004, India. bagariaivaibhav@gmail.com

Abstract

BACKGROUND

Deep learning, a form of artificial intelligence, has shown promising results for interpreting radiographs. In order to develop this niche machine learning (ML) program of interpreting orthopedic radiographs with accuracy, a project named deep learning algorithm for orthopedic radiographs was conceived. In the first phase, the diagnosis of knee osteoarthritis (KOA) as per the standard Kellgren-Lawrence (KL) scale in medical images was conducted using the deep learning algorithm for orthopedic radiographs.

AIM

To compare efficacy and accuracy of eight different transfer learning deep learning models for detecting the grade of KOA from a radiograph and identify the most appropriate ML-based model for the detecting grade of KOA.

METHODS

The study was performed on 2068 radiograph exams conducted at the Department of Orthopedic Surgery, Sir HN Reliance Hospital and Research Centre (Mumbai, India) during 2019-2021. Three orthopedic surgeons reviewed these independently, graded them for the severity of KOA as per the KL scale and settled disagreement through a consensus session. Eight models, namely ResNet50, VGG-16, InceptionV3, MobilnetV2, EfficientnetB7, DenseNet201, Xception and NasNetMobile, were used to evaluate the efficacy of ML in accurately classifying radiographs for KOA as per the KL scale. Out of the 2068 images, 70% were used initially to train the model, 10% were used subsequently to test the model, and 20% were used finally to determine the accuracy of and validate each model. The idea behind transfer learning for KOA grade image

classification is that if the existing models are already trained on a large and general dataset, these models will effectively serve as generic models to fulfill the study's objectives. Finally, in order to benchmark the efficacy, the results of the models were also compared to a first-year orthopedic trainee who independently classified these models according to the KL scale.

RESULTS

Our network yielded an overall high accuracy for detecting KOA, ranging from 54% to 93%. The most successful of these was the DenseNet model, with accuracy up to 93%; interestingly, it even outperformed the human first-year trainee who had an accuracy of 74%.

CONCLUSION

The study paves the way for extrapolating the learning using ML to develop an automated KOA classification tool and enable healthcare professionals with better decision-making.

Key Words: Osteoarthritis; Artificial intelligence; Knee; Computer vision; Machine learning; Deep learning

©The Author(s) 2022. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: In this study, we evaluated different machine learning models to determine which model is best to classify the severity of knee osteoarthritis using the Kellgren-Lawrence grading system. The image set was composed of radiographs of native knees, in anteroposterior and lateral views. The radiographic exams were annotated by experts and tagged according to Kellgren-Lawrence grades. The findings of this study will pave the way for future development in the field, with the development of more accurate models and tools that can improve medical image classification by machine learning and will give valuable insight into orthopedic disease pathology.

Citation: Tiwari A, Poduval M, Bagaria V. Evaluation of artificial intelligence models for osteoarthritis of the knee using deep learning algorithms for orthopedic radiographs. *World J Orthop* 2022; 13(6): 603-614

URL: <https://www.wjgnet.com/2218-5836/full/v13/i6/603.htm>

DOI: <https://dx.doi.org/10.5312/wjo.v13.i6.603>

INTRODUCTION

Knee osteoarthritis (KOA) is a debilitating joint disorder that degrades the knee articular cartilage and is characterized by joint pain, stiffness, and deformity, and pathologically by cartilage wear and osteophyte formation. KOA has a high incidence among the elderly, obese, and sedentary. Early diagnosis is essential for clinical treatments and pathology[1,2]. The severity of the disease and the nature of bone deformation and intra-articular pathology impact decision-making. Radiographs form an essential element of preoperative planning and surgical execution.

Artificial intelligence (AI) is impacting almost every sphere of our lives today. We are pushing the boundaries of human capabilities with unmatched computing and analytical power. Newer technological advancements and recent progress of medical image analysis using deep learning (DL), a form of AI, has shown promising ways to improve the interpretation of orthopedic radiographs. Traditional machine learning (ML) has often put much effort into extracting features before training the algorithms. In contrast, DL algorithms learn the features from the data itself, leading to improvisation in practically every subsequent step. This has turned out to be a hugely successful approach and opens up new ways for experts in ML to implement their research and applications, such as medical image analysis, by shifting feature engineering from humans to computers.

DL can help radiologists and orthopedic surgeons with the automatic interpretation of medical images, potentially improving diagnostic accuracy and speed. Human error due to fatigue and inexperience, which causes strain on medical professionals, is overcome by DL, reducing their workload and, most importantly, providing objectivity to clinical assessment and decisions for need of surgical intervention. Moreover, DL methods trained based on the expertise of senior radiologists and orthopedic surgeons in major tertiary care centers could transfer that experience to smaller institutions and create more space in emergency care, where expert medical professionals might not be readily available. This could dramatically enhance access to care. DL has been successfully used in different orthopedic applications, such as fracture detection[3-6], bone tumor diagnosis[7], detecting hip implant mechanical loosening[8], and grading OA[9,10]. However, the implementation of DL in orthopedics is not without challenges and limitations.

While most of our orthopedic technological research remains focused on core areas like implant longevity and biomechanics, it is essential that orthopedic teams also invest their effort in an exciting area of research that is likely to have a long-lasting impact on the field of orthopedics. With this aim, we set up a project to determine the feasibility and efficacy of the DL algorithms for orthopedic radiographs (DLAOR). To the best of our knowledge, only few published studies have applied DL to KOA classification, but many of these involved application to preprocessed, highly optimized images[11]. In this study, radiographs from the Osteoarthritis Initiative staged by a team of radiologists using the Kellgren-Lawrence (KL) system were used. They were standardized and augmented automatically before using the images as input to a convolution neural network model.

The purpose of this study was to compare eight different transfer learning DL models for KOA grade detection from a radiograph and to identify the most appropriate model for detecting KOA grade.

MATERIALS AND METHODS

The study was approved by an institutional review committee as well as the institutional ethics committee (HNH/IEC/2021/OCS/ORTH 56). It involved a diagnostic method based on retrospectively collected radiographic examinations. The examinations had been performed by a neural network, for assessment of both presence and severity of KOA using the KL system[12]. AI identifies patterns in images based on input data and associated recurring learning. It is fed with both the input (the radiographic images) and the information of expected output label (classifications of OA grade) to establish a connection between the features of the different stages of KOA (*e.g.*, possible osteophytes, joint space narrowing, sclerosis of subchondral bone) and the corresponding grade[13]. Before being fed to the classification engine, images were manually annotated by a team of radiologists according to the KL grading.

Data group

All the medical images were de-identified, without personal details of patients. The inclusion and exclusion criteria were applied after de-identification, complying with international standards like the Health Insurance Portability and Accountability Act.

Inclusion criteria

The following inclusion criteria were used: age above 18 years; recorded patient complaint of chronic knee pain (pain most of the time for more than 3 mo) and presentation to the orthopedic department or following radiologic investigation; and recorded patient unilateral knee complaints and knee grading according to KL system classification.

Exclusion criteria

The following exclusion criteria were used: patients who had undergone operation for total knee arthroplasty in either of the knees; patients who had undergone operation for unicompartmental knee arthroplasty; or patients who had recorded bilateral knee complaints.

Standardized knee X-rays: For the anteroposterior view of the knee, the leg had been extended and centered to the central ray. The leg had then been rotated slightly inward, in order to place the knee and the lower leg in a true anterior-posterior position. The image receptor had then been centered on the central ray. For a lateral view, the patient had been positioned on the affected side, with the knee flexed 20-30 degrees.

Assessment by orthopedic team: Three orthopedic surgeons reviewed these independently, graded them for the severity of KOA as per the KL scale and settled exams with disagreement through a consensus session. Finally, in order to benchmark the efficacy, the results of the models were also compared to a first-year orthopedic trainee who independently classified these models according to the KL scale.

Pre-processing

The models were trained with 2068 X-ray images. The dataset was split into three subsets used for training, testing, and validation, with a split ratio of 70-10-20. Initially, 70% of the images were used to train the model. Subsequently, 20% were used to test, and finally 10% were used for validation of the models. An end-to-end interpretable model took full knee radiographs as input and assessed KOA severity according to different grades (Figure 1).

Input images

The algorithm processed the images collected for each class. Each DICOM format radiograph was automatically cropped to the active image area, *i.e.* any black border was removed and reduced to a maximum of 224 pixels. Image dimensions were retained by padding the rectangular image to a square

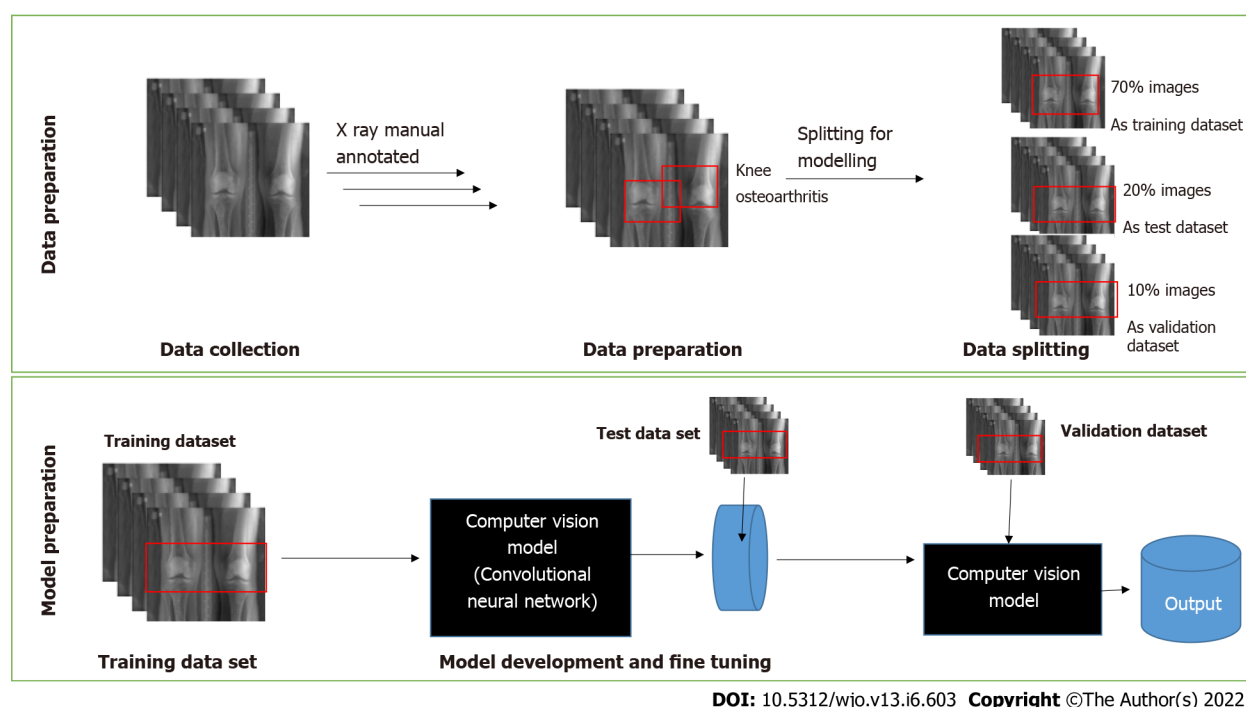


Figure 1 Architecture for deep learning algorithms for orthopedic radiographs.

format of 224×224 pixels. The training was carried out by cropping and rotating. Figure 2 shows the labelling of the KOA images in different classes as 0, 1, 2, 3 and 4, as per the KL grading scale and the knee severity from the X-ray images. The data were split into five different classes as per the KL grade shown in Figure 2. Table 1 shows the data splits in the training, testing and validation subsets, according to KL grades.

Need of transfer learning

A DL model is a program that has been trained on a set of data (called the training set) to recognize certain types of patterns. AI models use different algorithms to understand the information as part of a dataset and then learn from these data, with the definite goal of solving business challenges and overarching problems. The focus of traditional ML techniques used in image classification is acquiring knowledge from existing data and classifying. Most commonly, this means synthesizing practical concepts from historical data. However, another technique that encapsulates knowledge learned across many tasks and transfers it to new, unseen ones is transfer learning. Knowledge transfer can help speed up training and prevent overfitting, and improve the obtainable final performance. In transfer learning, knowledge is transferred from a trained model (or a set thereof) to a new model by encouraging the new model to have similar parameters. The trained models from which knowledge is transferred is not trained with this transfer in mind, and hence the task it was trained on must be very general for it to encode useful knowledge concerning other tasks. Transfer learning appeared to be helpful when there was a small training dataset with similar feature images. Thus, we have opted for a transfer learning approach to develop the DLAOR system. Figure 3 shows the traditional ML *vs* transfer learning.

DLAOR system

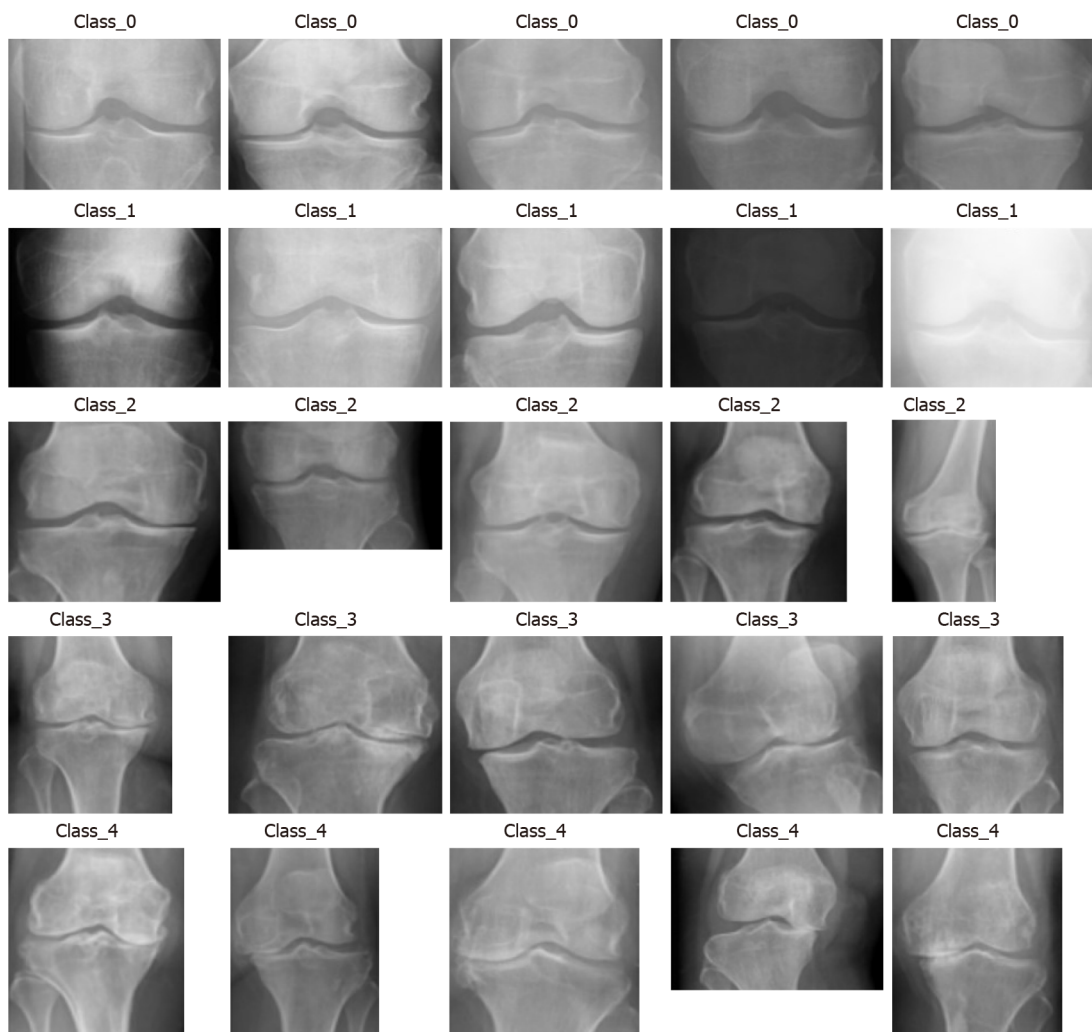
DLAOR is an image classification system where images are fed into the model to classify the KOA grade in medical images. It is built using a data ingestion pipeline, modelling engine, and classification system. The present study utilized a transfer learning technique for the effective classification of KOA grades using medical images. Since there are multiple well-established transfer learning models, we decided to use eight classification models. The selected models were reported to be the most efficient for image classification. All the selected models were compared to identify the best technique for the DLAOR system. The same set of images had been given to two OA surgeons to classify KL grades based on expert human interpretation and compared with models as part of the DLAOR system (Figure 4).

Model development

In this study, transfer ML models were developed using Python as Language; the Google Colab platform was used to run the code using GPU configuration to process models faster. Model development has the following steps: (1) Data augmentation and generators. There is the development of augmentation functions and data generator functions for training, testing, and validation datasets in

Table 1 Data splits in training, testing and validation subsets according to Kellgren-Lawrence grades

Osteoarthritis Kellgren-Lawrence grade	Training		Testing		Validation	
	Samples, <i>n</i>	Proportion, %	Samples, <i>n</i>	Proportion, %	Samples, <i>n</i>	Proportion, %
0	255	17.6	37	17.6	73	17.7
1	213	14.7	31	14.7	61	14.8
2	164	11.4	24	11.4	47	11.4
3	237	16.4	35	16.7	68	16.4
4	576	39.9	83	39.5	164	39.7
Total	1445	100	210	100	413	100



DOI: 10.5312/wjo.v13.i6.603 Copyright ©The Author(s) 2022.

Figure 2 X-ray images of different Kellgren-Lawrence grades for knee osteoarthritis.

this stage; (2) Import the base transfer learning model. The next step is to import the base model from the TensorFlow/Keras library; (3) Build and compile the model. The base model was used just as a layer in a sequential model. After that, a single fully connected layer on top was added. Then, the model was compiled with Adam as an optimizer and categorical cross entropy as a loss mechanism; (4) Fit the model. In this step, the final model was fitted over the validation dataset. Then, with the help of 10 epochs, the model was finalized. This model was then evaluated over the test data; and (5) Fine-tune the model. We fine-tuned the model to improve the statistical order by giving a learning rate of 0.001 and additional epochs as 20.

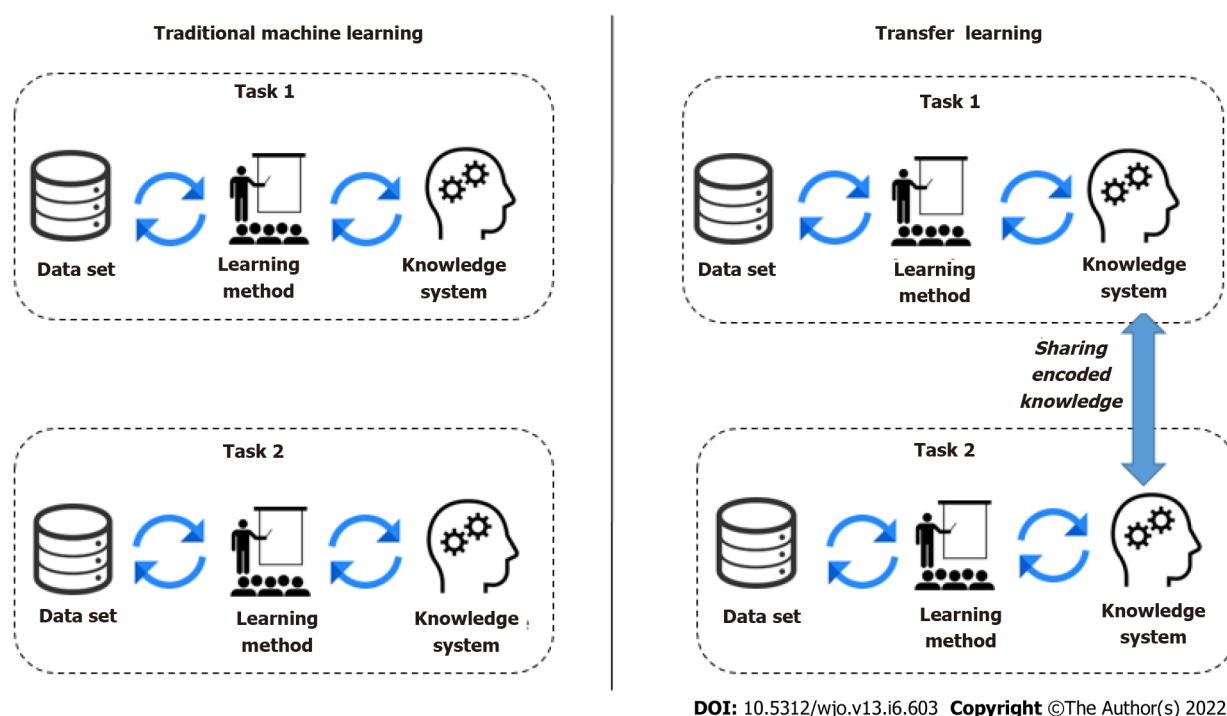


Figure 3 Traditional machine learning vs transfer learning.

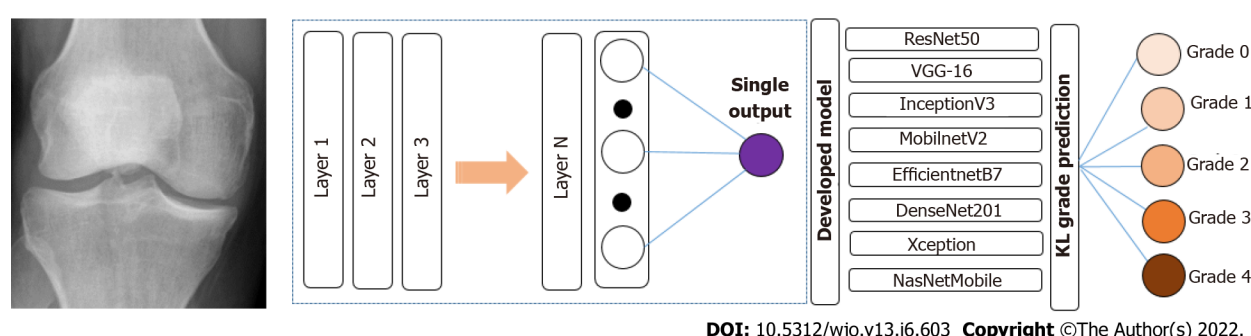


Figure 4 Multimodal pipeline, predicting the risk of osteoarthritis for a particular knee. We first used a deep convolutional neural network, trained different models in a multitask setting to predict the current stage of osteoarthritis defined according to the Kellgren-Lawrence (KL) grade scale.

RESULTS

In the present study, transfer ML was performed to compare the following parameters, to compare which model or human expert interpretation performed better regarding image classification of KOA grades: (1) Accuracy, which tells us what fractions are labeled correctly; (2) Precision, which is the number of true positives (TPs) over the sum of TPs and false positives. It shows what fraction of positive sides are correct; (3) Recall, which is the number of TPs over the sum of TPs and false negatives. It shows what fraction of actual positives is correct; and (4) Loss, which indicates the prediction error of the model.

Table 2 shows the performance comparison of the various models used for transfer learning models for KOA with the accuracy, precision, recall, and loss. It was observed that the highest accuracy was 92.87% and was achieved using DenseNet201. It had high precision and recall value as compared to all the other models.

This study leveraged cross-entropy multiple class loss. It is the measure of divergence of the predicted class from the actual class. It was calculated as a separate loss for each class label per observation and sums the result (Table 3).

Figure 5 shows the impact of iterations on the loss and accuracy of the eight different models. All the image clarification models had accuracy over 54%, while five models achieved accuracy over 80%, and one model achieved accuracy above 90%. High precision was achieved as 93.69% from the DenseNet model. The same model resulted in the highest accuracy of 92.87% and recall of 92.53%. Maximum loss was observed for the VGG-16 model. The minimum value for accuracy and precision was observed in

Table 2 Evaluation of parameters for knee osteoarthritis detection

Parameter	
Accuracy	Determines the accuracy of the standalone model inaccuracy to detect the presence of KOA and its classification in the input image
Precision	True positive/true positive + false positive
Recall	True positive/true positive + false negative
Loss	Determines the loss of the model

KOA: Knee osteoarthritis.

Table 3 Performance comparison of various transfer learning convolutional neural network models and eight expert human interpretations used for the development of deep learning algorithm for orthopedic radiographs

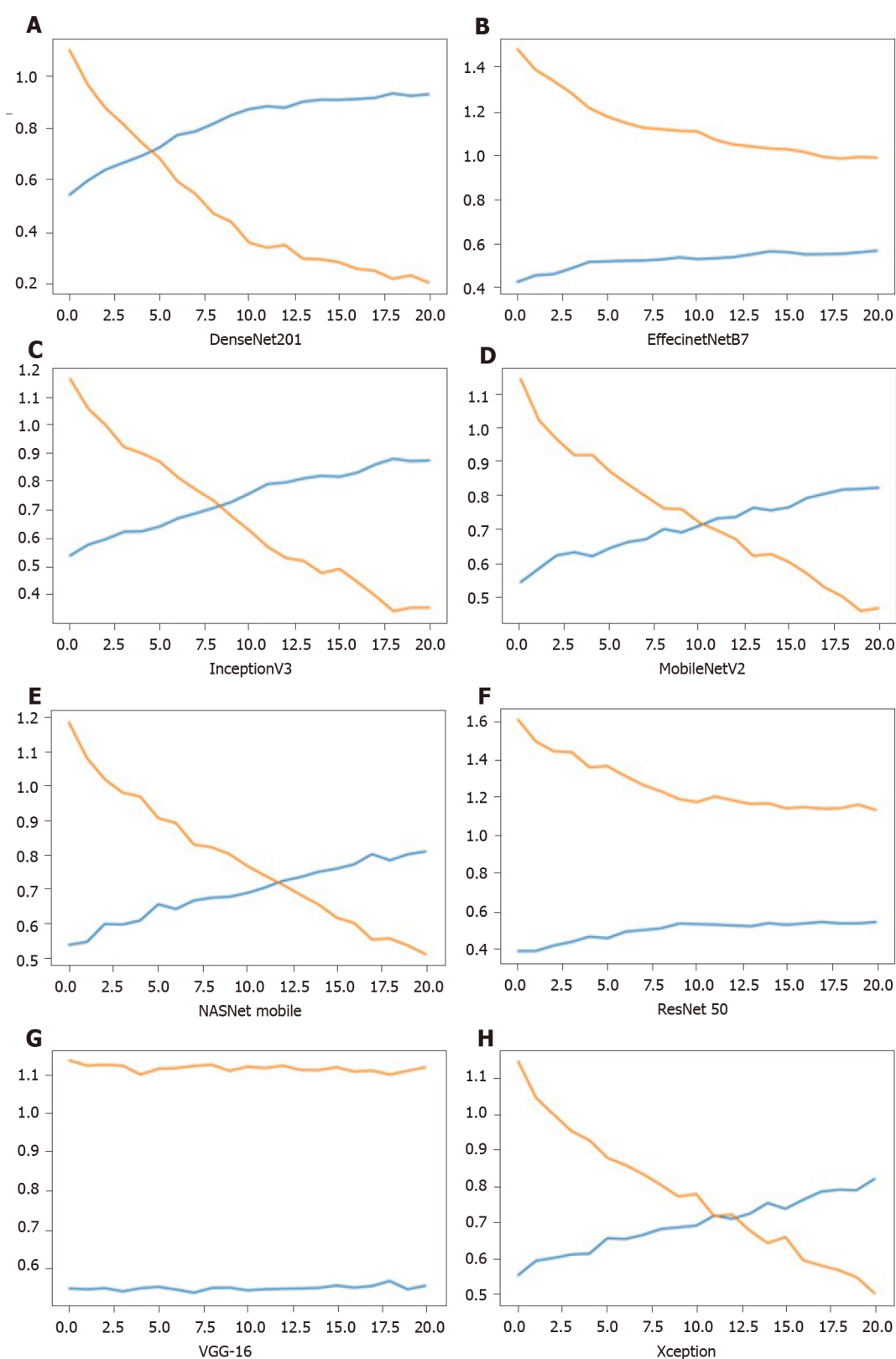
Model name	Accuracy	Precision	Recall	Loss	Outcome
ResNet50	54.29%	61.03%	39.52%	1.06	Average
VGG-16	56.68%	67.56%	35.02%	1.10	Average
InceptionV3	87.34%	89.19%	85.67%	0.35	Good
MobilenetV2	82.15%	84.66%	80.21%	0.46	Average
EfficientnetB7	56.61%	70.09%	38.27%	0.98	Average
DenseNet201	92.87%	93.69%	92.53%	0.20	Best
Xception	82.81%	85.03%	77.05%	0.50	Average
NasNetMobile	80.90%	83.98%	77.30%	0.50	Average
Surgeon	74.22%	79.50%	50.00%	0.25	Good

the ResNet50 model. The result showed that DenseNet was the best model to develop a DLAOR image classification system. The expert human interpretation for image classification was also captured for comparison. This was performed by two OA surgeons, and observed accuracy was 74.22%, with minimum loss value as second in comparison to all models.

DISCUSSION

Making the DL algorithm interpretable can benefit both the fields of computer science (to enable the development of improved AI algorithms) and medicine (to enable new medical discoveries). Interpreting and understanding cases of a machine's failure or when the machine is underperforming relative to the human expert can improve the DL algorithm itself. Understanding the AI decision-making process when the machine is outperforming the human expert can lead to new medical discoveries to extract the deep knowledge and insights that the AI utilizes to more optimal decisions than humans. DL in healthcare has already achieved wonders and improved the quality of treatment. Multiple organizations like Google, IBM, *etc* have spent a significant amount of time developing DL models that make predictions around hospitalized patients and support managing patient data and results. The future of healthcare has never been more intriguing, and both AI and ML have opened opportunities to craft solutions for particular needs within healthcare.

Standard radiographs of the hand, wrist, and ankle fractures were automatically diagnosed as a first step while identifying the examined view and body part[5]. A continuous, reliable, fast, effective, and most importantly accurate machine would be an asset without access to a radiology expert in the premise for preliminary diagnosis. AI solutions may incorporate both image data and the radiology text report for best judgment of an image or further learning of the network[14]. A recent study[15] revealed that with the help of isolating the subtle changes in cartilage texture of knee cartilage in a large number of magnetic resonance imaging T2-mapped images could effectively predict the onset of early symptomatic (WOMAC score) KOA, which may occur 3 years later. We are in a new era of orthopedic diagnostics where computers and not the human eye will comprehend the meaning of image data[16, 17], creating a paradigm shift for orthopedic surgeons and even more so for radiologists. The application of AI in orthopedics has mainly focused on the implementation of DL on clinical images.



DOI: 10.5312/wjo.v13.i6.603 Copyright ©The Author(s) 2022.

Figure 5 Depicting the loss and accuracy vs number of epochs for different models. Red line: Loss; Blue line: Accuracy; Y-axis: Depicting the loss and accuracy; X-axis: Number of epochs. A: DenseNet201; B: EffecinetNetB7; C: InceptionV3; D: MobileNetV2; E: NasNetMobile; F: ResNet 50; G: VGG-16; H: Xception.

DL algorithms will diagnose symptomatic OA needing treatment from radiological OA with no symptoms needing just observation. Both humans and machines make mistakes in applying their intelligence to solving problems. In ML, overfitting memorizes all examples, and an overfitted model lacks generalization and fails to work on never-before seen examples. In ML, transfer learning is a

method that reuses a model that has been developed through a system getting to know specialists and has already been reinforced on a huge dataset[18,19].

Transfer learning leverages data extracted from one set of distributions and uses it to leverage its learning to predict or classify others. In humans, the switch of expertise to college students is frequently performed with the aid of using instructors and training providers. This might not make the scholars smart. However, transfer learning makes the knowledge provider wise because of the knowledge-gathering party. In the case of humans, transferring expertise relies upon the inherent intelligence of the expert to improve the classification and prediction skills.

Future prospects

With the identification of top models for the classification of OA using a medical image, these models can be deployed for automating the process of preliminary X ray processing and determination of the next course of action. Any image classification project starts with a substantial number of images classified by an expert. This will provide inputs to the training of the model. There are multiple challenges that can result from this misclassification of data during validation of the model by a machine, such as intra-class variation, image scale variation, viewpoint variation, *i.e.* orientation of image, occlusion, multiple objects in the image, illumination, and background clutter. While humans can overcome some of these possible challenges with naked eyes, it will be difficult for a machine to reduce the impacts of such factors while classifying the images. This results in an increased amount of loss and effect on the outcome of the prediction.

Expert human interpretation accuracy comes with experience, and it was observed in this study. Both humans and machines are prone to overfitting, a curse that impacts preconceived notions while making decisions. AI is limited to the areas in which they are trained. Human intelligence and its interpretation, however, are independent of this domain of training. The key differentiator between human beings and machines is that humans will be able to solve problems related to unforeseen domains, while the latter will not have the capability to do that. The common solution suggested by experts to overcome overfitting or underfitting is cross-validation and regularization. This would be achieved by increasing the size or number of parameters in the ML model, increasing the exploration of complexity or type of the model, increasing the training time, and increasing the number of iterations until loss function in ML is minimized. Another major drawback of utilizing transfer learning models for image classification is negative transfer, where transfer learning ends up decreasing the accuracy or performance of the model. Thus, there is a need to customize the models and their parameters to see the impact of tuning and epochs on the performance or accuracy of the model.

This study is a preliminary step to the implementation of an end-to-end machine recommendation engine for the orthopedic department, which can provide sufficient inputs for OA surgeons to determine the requirement of OA surgery, patient treatment, and prioritization of patients based on engine output and related patient demographics.

CONCLUSION

The present study found that a neural network can be trained to classify and detect KOA severity and laterality as per the KL grading scale. The study reported excellent accuracy with the DenseNet model for the detection of the severity of KOA as per KL grading. Previously, it was challenging for healthcare professionals to collect and analyze large data for classification, analysis, and prediction of treatments since there were no such tools and techniques present. But now, with the advent of ML, it has been relatively available. ML may also help provide vital statistics, real-time data, and advanced analytics in terms of the patient's disease, lab test results, blood pressure, family history, clinical trial data, *etc* to doctors. The future of the technology in orthopedics and the application of this study's findings are multi-fold. We can also borrow from experience from other subspecialties.

Transfer learning techniques will not work when the image features learned by the classification layer, which are not enough nor appropriate to identify the classes for KL grades. When the medical images are not similar, then the KOA features are transferred poorly in the model as in the conventional model building, where layers for accuracy purposes have been adjusted as per convenience. If the same is applied in transfer learning, then it results in a reduction of trainable parameters and results in overfitting. Thus, it is important that a balance between overfitting and learning features is achieved.

This study also plans to improve the accuracy, which can be improved with more images, and the features of the algorithm. The implication of this study goes beyond the simple classification of KOA. The study paves the way for extrapolating the learning from KOA classification using ML to develop an automated KOA classification tool and enable healthcare professionals with better decision-making. This study helps in exploring trends and innovations and improving the efficiency of existing research and clinical trials associated with AI-based medical image classification. The Healthcare sector, including orthopedics, generates a large amount of real-time and non-real-time data. However, the challenge is to develop a pipeline that can collect these data effectively and use them for classification, analysis, prediction, and treatment in the most optimal manner. ML allows building models to quickly

analyze data and deliver results, leveraging historical and real-time data. Healthcare personnel can make better decisions on patients' diagnoses and treatment options using AI solutions, which leads to enhanced and effective healthcare services.

ARTICLE HIGHLIGHTS

Research background

Artificial intelligence (AI)-based on deep learning (DL) has demonstrated promising results for the interpretation of radiographs. To develop a machine learning (ML) program capable of interpreting orthopedic radiographs with accuracy, a project called DL algorithm for orthopedic radiographs was initiated. It was used to diagnose knee osteoarthritis (KOA) using Kellgren-Lawrence scales in the first phase.

Research motivation

By using DL methods trained by senior radiologists and orthopedic surgeons in larger hospitals, smaller institutions could gain the expertise they need and create more space in emergency care, where medical professionals may not be readily available. Providing care in this manner would improve access dramatically.

Research objectives

This study aimed to explore the use of transfer learning convolutional neural network for medical image classification applications using KOA as a clinical scenario, comparing eight different transfer learning DL models for detecting the grade of KOA from a radiograph, to compare the accuracy between results from AI models and expert human interpretation, and to identify the most appropriate model for detecting the grade of KOA.

Research methods

As per the Kellgren-Lawrence scale, three orthopedic surgeons reviewed these independent cases, graded their severity for OA, and settled disagreements through consensus. To assess the efficacy of ML in accurately classifying radiographs for KOA, eight models were used, including ResNet50, VGG-16, InceptionV3, MobilnetV2, EfficientNetB7, DenseNet201, Xception, and NasNetMobile. A total of 2068 images were used, of which 70% were used initially to train the model, 10% were then used to test the model, and 20% were used for accuracy testing and validation of each model.

Research results

Overall, our network showed a high degree of accuracy for detecting KOA, ranging from 54% to 93%. Some networks were highly accurate, but few had an efficiency of more than 50%. The DenseNet model was the most accurate, at 93%, while expert human interpretation indicated accuracy of 74%.

Research conclusions

The study has compared the accuracy provided by expert human interpretation and AI models. It showed that an AI model can successfully classify and differentiate the knee X-ray image with the presence of different grades of KOA or by using various transfer learning convolution neural network models against human actions to classify the same. The purpose of the study was to pave the way for the development of more accurate models and tools, which can improve the classification of medical images by ML and provide insight into orthopedic disease pathology.

Research perspectives

AI can only operate within the areas in which it has been trained, whereas human intelligence and its interpretation are independent of the area in which it has been trained. One of the key differences between humans and machines is that humans will be able to solve problems related to unforeseen domains, while the latter will not have the capability to do that. It can be accomplished by increasing the size or number of parameters in the ML model, examining the complexity or type of the model, increasing the time spent training, and increasing the number of iterations until the loss function in ML is minimized.

FOOTNOTES

Author contributions: Tiwari A analyzed and evaluated the feasibility and accuracy of the artificial intelligence models with the interpretation of data; Poduval M was involved in critically revising the draft; Bagaria V designed and conceptualized the study and revised the manuscript for important intellectual content; All authors read and

approved the final manuscript.

Institutional review board statement: The study was reviewed and approved by the Scientific Advisory Committee and Institutional Ethics Committee of Sir H. N. Reliance Foundation Hospital and Research Centre, Mumbai, India (Approval No. HNH/IEC/2021/OCS/ORTH/56).

Informed consent statement: Informed consent was waived since the data were retrospectively collated anonymously from routine clinical practice.

Conflict-of-interest statement: The authors of this manuscript have no conflicts of interest to disclose.

Data sharing statement: No additional data are available.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <https://creativecommons.org/licenses/by-nc/4.0/>

Country/Territory of origin: India

ORCID number: Anjali Tiwari 0000-0002-8683-6102; Murali Poduval 0000-0002-6821-0640; Vaibhav Bagaria 0000-0002-3009-3485.

S-Editor: Wang JL

L-Editor: A

P-Editor: Wang JL

REFERENCES

- 1 Hsu H, Siwiec RM. Knee Arthroplasty. In: StatPearls [Internet]. Treasure Island: StatPearls Publishing, 2022 [PMID: 29939691]
- 2 Aweid O, Haider Z, Saed A, Kalairajah Y. Treatment modalities for hip and knee osteoarthritis: A systematic review of safety. *J Orthop Surg (Hong Kong)* 2018; **26**: 2309499018808669 [PMID: 30415598 DOI: 10.1177/2309499018808669]
- 3 Urakawa T, Tanaka Y, Goto S, Matsuzawa H, Watanabe K, Endo N. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. *Skeletal Radiol* 2019; **48**: 239-244 [PMID: 29955910 DOI: 10.1007/s00256-018-3016-3]
- 4 Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol* 2018; **73**: 439-445 [PMID: 29269036 DOI: 10.1016/j.crad.2017.11.015]
- 5 Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A, Sköldenberg O, Gordon M. Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthop* 2017; **88**: 581-586 [PMID: 28681679 DOI: 10.1080/17453674.2017.1344459]
- 6 Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, Hanel D, Gardner M, Gupta A, Hotchkiss R, Potter H. Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci U S A* 2018; **115**: 11591-11596 [PMID: 30348771 DOI: 10.1073/pnas.1806905115]
- 7 Do BH, Langlotz C, Beaulieu CF. Bone Tumor Diagnosis Using a Naïve Bayesian Model of Demographic and Radiographic Features. *J Digit Imaging* 2017; **30**: 640-647 [PMID: 28752323 DOI: 10.1007/s10278-017-0001-7]
- 8 Langhorn J, Borjali A, Hippensteel E, Nelson W, Raeymaekers B. Microtextured CoCrMo alloy for use in metal-on-polyethylene prosthetic joint bearings: multi-directional wear and corrosion measurements. *Tribol Int* 2018; **124**: 178-183 [PMID: 30778273 DOI: 10.1016/j.triboint.2018.04.007]
- 9 Xue Y, Zhang R, Deng Y, Chen K, Jiang T. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. *PLoS One* 2017; **12**: e0178992 [PMID: 28575070 DOI: 10.1371/journal.pone.0178992]
- 10 Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. *Sci Rep* 2018; **8**: 1727 [PMID: 29379060 DOI: 10.1038/s41598-018-20132-7]
- 11 Tiulpin A, Klein S, Bierma-Zeinstra SMA, Thevenot J, Rahtu E, Meurs JV, Oei EHG, Saarakkala S. Multimodal Machine Learning-based Knee Osteoarthritis Progression Prediction from Plain Radiographs and Clinical Data. *Sci Rep* 2019; **9**: 20038 [PMID: 31882803 DOI: 10.1038/s41598-019-56527-3]
- 12 Pai V, Knipe H. Kellgren and Lawrence system for classification of osteoarthritis. [cited 23 Sep 2021]. Available from: <https://radiopaedia.org/articles/27111>
- 13 Khan A, Sohail A, Zahoor U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. *Artif Intell Rev* 2020; **53**: 5455-5516 [DOI: 10.1007/s10462-020-09825-6]
- 14 van Ginneken B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiol Phys Technol* 2017; **10**: 23-32 [PMID: 28211015 DOI: 10.1007/s12194-017-0394-5]
- 15 Ashinsky BG, Bouhrara M, Coletta CE, Lehallier B, Urish KL, Lin PC, Goldberg IG, Spencer RG. Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the

- osteoarthritis initiative. *J Orthop Res* 2017; **35**: 2243-2250 [PMID: [28084653](#) DOI: [10.1002/jor.23519](#)]
- 16 **Poduval M**, Ghose A, Manchanda S, Bagaria V, Sinha A. Artificial Intelligence and Machine Learning: A New Disruptive Force in Orthopaedics. *Indian J Orthop* 2020; **54**: 109-122 [PMID: [32257027](#) DOI: [10.1007/s43465-019-00023-3](#)]
- 17 **Hamet P**, Tremblay J. Artificial intelligence in medicine. *Metabolism* 2017; **69S**: S36-S40 [PMID: [28126242](#) DOI: [10.1016/j.metabol.2017.01.011](#)]
- 18 **Radford A**, Wu J, Amodei D, Clark J, Brundage M, Sutskever I. Better language models and their implications. OpenAI, 2018b. Available from: <https://openai.com/blog/better-language-models>
- 19 **Devlin J**, Chang M W, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. 2018 Preprint. Available from: arXiv:1810.04805 [DOI: [10.48550/arXiv.1810.04805](#)]



Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA

Telephone: +1-925-3991568

E-mail: bpgoffice@wjgnet.com

Help Desk: <https://www.f6publishing.com/helpdesk>

<https://www.wjgnet.com>

