International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Transfer Learning Based Intelligent Diagnosis Support System for Knee Osteoarthritis

Chuan-Jun Su[1], Yi Li[2]
[1, 2]Department of Industrial Engineering and Management Yuan-Ze University, Taoyuan, Taiwan
Email address: iecjsu @ saturn.yzu.edu.tw, s1068905 @ mail.yzu.edu.tw

Abstract— *Recent studies have determined that osteoarthritis has become one of the most common diseases of the knee joint. As an irreversible disease, with the gradual wear and tear of articular cartilage, its condition will gradually deteriorate, which will bring inconvenience to patients' lives.*

As the main criterion of osteoarthritis, knee x-rays usually require professional judgment from doctors with rich experience. However, a large number of x-rays and small differences between different stages often cause a heavy burden on doctors and make mistakes in judgment.

With the rapid development in recent years, machine learning technology or artificial intelligence (AI) has become a common solution to medical problems. In this research, a transfer learning based medical diagnosis support system was built to reduce the burden on physicians. Furthermore, a friendly user interface was designed for medical staff to efficiently access the functionality of the proposed system. After comparing different algorithms with different frameworks, a transfer learning model with 96.98% accuracy built into the system.

Keywords— *Transfer learning, Osteoarthritis, Medical images.*

I. INTRODUCTION

A. Background and motivation

Osteoarthritis (OA) is one of the most common joint diseases. It is a chronic debilitating joint disease characterized by degenerative changes to the bones, cartilage, menisci, ligaments, and synovial tissue [1]. The American Rheumatism Association (ARA) defines osteoarthritis as a heterogeneous group of conditions that lead to joint symptoms and signs which are associated with defective integrity of cartilage, in addition to related changes in the underlying bone and at the joint margin [2].

Osteoarthritis ranks first among the 50 most common diseases and sequelae of injuries in the world, affecting over 250 million people, accounting for 4 % of the world’s population. It is related to age, obesity, old injuries of joints, occupation, and genes. The knee joint plays a very important role in daily walking, so once osteoarthritis occurs, it will be difficult to walk due to pain. To avoid this happening, it’s important to diagnose osteoarthritis before the osteoarthritis becomes serious.

Joint X-ray is the main method to diagnose osteoarthritis [3]. Currently, the Kellgren and Lawrence (KL) classification is the most widely used clinical tool for the radiographic diagnosis of osteoarthritis [1]. The KL classification is usually applied specifically within the context of knee osteoarthritis. Joint X-rays are assigned a grade from 0 to 4, which they correlated to the increasing severity of osteoarthritis, with Grade 0 signifying no presence of osteoarthritis and Grade 4 signifying severe osteoarthritis [4].

Fig. 1 shows the radiographs of osteoarthritis in each KL scale. It can be seen that the KL scale is graded according to the space between the bones and whether the osteophyte grows.

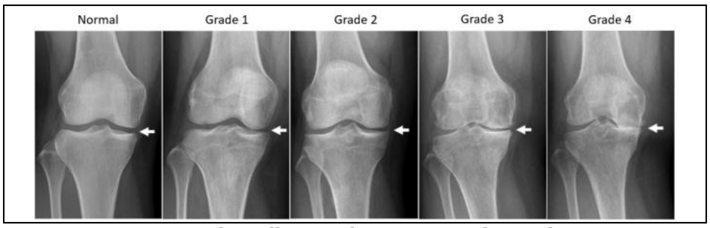

Fig. 1. The Kellgren and Lawrence grading scale

In recent years, with new technological and algorithm advances, artificial intelligence has become more popular. Many industries are slowly applying AI technology in their own fields, such as traditional industries, manufacturing, telecommunications, service, etc.

In addition, many major medical centers have actively invested in research on artificial intelligence to introduce artificial intelligence into the medical industry. It can lighten doctor’s load and reduce time to treat patients. It allows more patients to receive medical services.

B. Research Objective

This purpose of this study is to provide a diagnostic system that can identify which level osteoarthritis is. It can assist medical staff to diagnose the severity of knee arthritis to reduce judgment time. This study takes knee arthritis as an example. Use knee joint X-ray to establish a model for identifying the KL grade of knee arthritis.

C. Research Overview

The remainder of this research is organized as these following. Chapter 2 organizes the literature related to this research and explore related applications, such as deep learning, convolutional neural network, transfer learning, and some medical researches. Chapter 3 explains the research methodology for the knee arthritis grading system. Chapter 4 describes this system implementation details, such as model prediction status. Chapter 5 summarizes the conclusions and suggestions for future research.

Chuan-Jun Su and Yi Li, “Transfer Learning Based Intelligent Diagnosis Support System for Knee Osteoarthritis,” *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 3, Issue 8, pp. 62-67, 2021.

II. LITERATURE REVIEW

A. Degenerative Arthritis

According to the Osteoarthritis Research Society International (OARSI), Degenerative arthritis (Osteoarthritis) is an undesirable repair reaction caused by micro injury or macro-injury, including inflammation and innate immune pathways, which in turn causes the articular cartilage to gradually degenerate. This disease is originally due to abnormal tissue metabolism leading to joint changes, such as cartilage degradation, bone spurs, and joint inflammation, and finally lost normal joint function [5].

Degenerative arthritis can be classified into primary degenerative arthritis and secondary degenerative arthritis. Primary degenerative arthritis is caused by age and gender. It occurs more commonly in men before the age of 45 and women between the ages of 45 and 55, and the symptoms are more severe in women after menopause. This disease may be closely related to changes in the secretion of estrogen in the body. The symptoms are most obvious in people between the age of 55 and 65. According to research, almost every 70-year-old person has a certain of osteoarthritis, and women are more likely to suffer from osteoarthritis than men [6]. And secondary degenerative arthritis is usually caused by injuries, obesity, metabolic diseases, systemic diseases, and congenital abnormalities.

The most common joints of degenerative arthritis are the hip, knee joints, and other important parts that support the weight of the whole body. It may also occur in any joint position of the body. The spine and hand joints may also be affected, and the most common part of the hand is the distal knuckles. The difference between degenerative arthritis and rheumatoid arthritis is that this disease is a local, non-inflammatory joint disease. When people are young, cartilage repairability is better, but after decades of wear and tear, coupled with the degeneration phenomenon of aging, the cartilage repair will become slow, the cartilage becomes thinner and thinner, and even the cartilage is eroded and destroyed [7].

B. Deep Learning

In recent years, as technology advances, computers become more powerful, and many algorithms have been proposed, so artificial intelligence develops rapidly. In addition, it has become popular, and many companies have actively invested in deep learning research, such as Google and Amazon. There are two more well-known artificial intelligence technologies, namely machine learning and deep learning, especially deep learning.

Deep learning is a subset of machine learning. It is an algorithm for learning how to analyze data characteristics based on a neural network. Previously, the computer's computing power was insufficient and there were no enough training data, so the development of deep learning was limited. Now, as the GPU has improved the computing power of the computer and many sensors have captured a lot of data, it has developed rapidly. It consists of three layers, namely the input layer, the hidden layer, and the output layer. The architecture of the deep learning model is layered to connect the layers.

Deep learning has made many achievements in many fields, such as predictive maintenance, computer vision, and natural language processing. For example, it performs well in face recognition [8], object detection [9,10], speech recognition [11], and scene reconstruction [12].

There are a lot of deep learning models. The typical models include Artificial Neural Networks (ANN), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN). CNN is good at dealing with image data. RNN is adept at processing time-series data.

In this study, our input is knee joint X-ray images, so we determine to build the model with CNN. This CNN model can assist a doctor in diagnosing the degree of osteoarthritis.

C. Convolutional Neural Network

Convolutional neural network is one of the most representative neural networks in the field of deep learning. It is good at extracting image features and analyzing what the image is. A well-known application in convolutional neural network is handwriting recognition that uses the mnist dataset [13,14]. Convolutional neural network consists of three layers combining convolution layer, pooling layer, and fully connected layer.

D. Transfer Learning

Transfer learning is a branch of machine learning. It takes the old solution model and uses it on other different but related new problems. For example, the model used to identify cats and dogs can also be used to improve the ability to identify other animals. In some cases, we don't have enough information to build a model. In these cases, because transfer learning is feasible, we can try to use other models to see if good performance can also be obtained.

With new technological advances, artificial intelligence has become more popular. More and more people have begun to study machine learning, and many algorithms and architectures of the model have also been published. There are also many classic architectures of the CNN model released publicly. In transfer learning, we can use these famous models to solve our problems. So far, the famous models that have been proposed are AlexNet, VGGNet, GoogLeNet, ResNet, DenseNet, MobileNet, and EfficientNet.

E. Medical Research

With the advancement of artificial intelligence technology, the medical industry has begun to integrate with artificial intelligence. The US Food and Drug Administration has also approved some artificial intelligence-based medical devices, such as IDx-DR [15]. There are also some related articles published, such as the use of deep learning to diagnose chest X-rays[16] and knee joint X-rays [17].

III. RESEARCH METHODOLOGY

The main purpose of medical images is to understand the images of the human body in a non-invasive way during medical treatment or medical research. After the German physicist Wilhelm Conrad Rontgen discovered X-rays in 1895, X-rays became one of the most commonly used types of medical images. Physicians can use X-rays to understand the internal body of patients to determine the patient's physical health.

In recent years, because of the rapid rise of AI, there have been many cases of deep learning combined with medical images. The first application that appeared in 2016 was a paper published by GOOGLE's scientific research organization. It used deep learning to detect diabetic retinopathy early. It detects the disease early to prevent blindness through the image of the retina of diabetic patients. This shows that we may be able to use deep learning to analyze the X-ray image of the knee joint, find out the features, and analyze the grade to which it belongs, thereby assisting the medical staff in the diagnosis process.

This chapter introduces the research architecture of this study, which uses deep learning to diagnosis the grade of the osteoarthritis system for assisting medical staff in their work.

A. *The Overview of Research Architecture*

This research can be divided into three stages, the first stage is to collect the knee X-rays, the second stage is to pre-process the knee X-rays, and the last stage is to train the predicted model. First, we collaborate with the hospital to collect the knee X-rays. Next, we asked the doctor to help classify the knee X-rays. Finally, we trained a model that can diagnose the K/L grade of knee arthritis through the knee X-rays.

B. *Knee X-Rays Data*

Thanks to the cooperation with the hospital, the knee X-rays used in this study were obtained. There are a total of 1453 X-rays, which can be divided into five grades and unrecognizable X-ray images. 0 is normal Knee X-rays images, 1-4 is the severity of arthritis, and na is an unrecognizable X-ray image that means the joint is an artificial joint, so it cannot be determined. There are 62 X-rays images at grade 0, 149 X-rays images at grade 1, 342 X-rays images at grade 2, 411 X-rays images at grade 3, 431 X-rays images at grade 4, and 58 X-rays images of artificial joints.

C. *Labeling and Pre-processing*

In the beginning, the X-ray images are collected first, and then ask medical staff to help label the X-ray images, and divide the X-ray images into five K/L grades and artificial joints. We can find six types of folders in the osteoarthritis folder. After the classification, since the main judgment of degenerative arthritis is the condition of the joints, we used Labelimg which is a labeling image tool to label the X-ray joint, and then cut the range of the label by using python, capture the part we need. Fig. 2 shows the process of cropping knee X-ray images.

D. *Training Model*

After collecting the knee X-ray images and the pre-processing of the knee X-ray images, the next step is to train the model to judge the grade of the knee X-ray images. Since the research in this article is to analyze images, we use the convolutional neural network in deep learning as the basis for training the model. The first step is to set the image size of the input layer. The second step is to set the convolutional layer and the pooling layer to capture the features of the images, we will use the architecture of Inception V3, ResNet, DenseNet, MobileNet, and EfficientNet to build the model. The last step is to set the fully connected layer to analyze which K/L grade the knee X-ray image is.

Fig. 2. The process of preprocessing knee X-ray images

In this research, we use Keras to build the model. Keras is an open-source high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. Using Keras just concentrates on building models, so it uses the least amount of code and spends the least time to build a deep learning model, train them, evaluate the accuracy, and make predictions.

E. *Designing GUI*

After training the model, we designed a user interface for doctors to operate, so that doctors can use the diagnostic model to diagnose the knee X-ray images. With this GUI, doctors can easily use the diagnostic model to assist them in diagnosing the knee X-ray images.

The user interface is mainly divided into two blocks. The first block is to select the knee X-ray image to be diagnosed, and the system will automatically use the diagnostic model to diagnose the knee X-ray image, and the second block is to display the diagnosis result to the user.

IV. IMPLEMENTATION

A. *Image Data Processing Phase*

Before starting to build a diagnostic model, we need to make some preparations. First, we need to collect knee X-rays for training. Then we need to ask the doctor to label the knee X-ray according to the Kellgren-Lawrence classification. Finally, we need to pre-process the knee X-rays. Since the model only needs the joints of the knee joint to be diagnosed, the rest is not important, so we decided to cut the knee X-rays, leaving only the joints of the knee X-rays. LabelImg is a tool we use to label images. Use LabelImg to mark the position of the knee joint first, and it will store the marked position in XML format. Fig. 3 shows the process of marking knee X-ray images.

Next, in order to cut all the knee X-rays quickly, we write python to use the coordinates in the XML file to cut the X-rays of the knee joints, leaving the joints of the knee joints. After that, the cropped knee joint X-rays can be used for model training. Fig. 4 shows the process of cropping knee X-ray images.

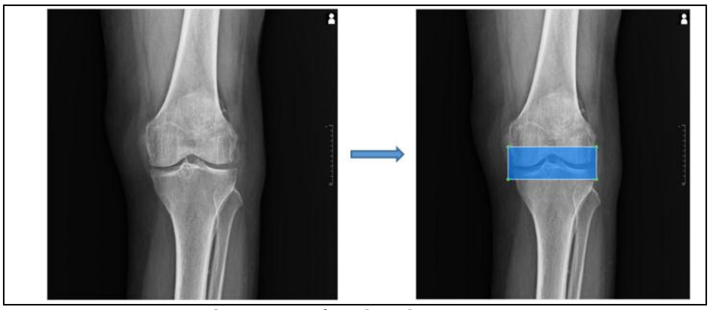

Fig. 3. The process of marking knee X-ray images

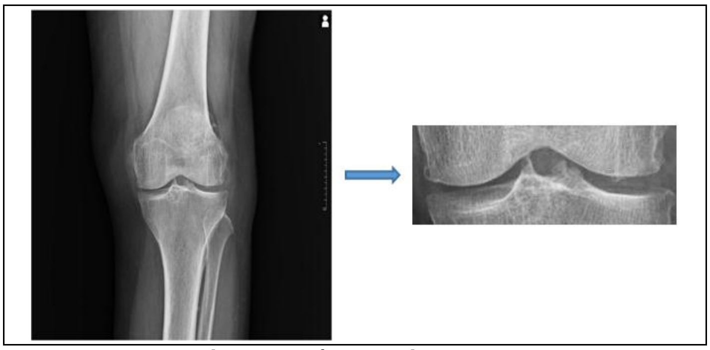

Fig. 4. The process of cropping knee X-ray images

B. Model building phase

After we preprocessed our data set, we can start to build a diagnosis model for osteoarthritis. This research uses transfer learning to construct the model. We use five models for comparison. These five models are Inception V3, ResNet, DenseNet, MobileNet, and EfficientNet. They are all well-known neural network models, and they all have good rankings in the ILSVRC classification competition.

Accuracy is commonly the most important metric in AI-based systems, while recall and precision provide more information for further accessing the performance of the system. In this research, type I error is more serious than type II error. Therefore, when accuracy is the same, we will use recall to decide which model will be used. Then, we divide the diagnosis process into three stages, so there will train three models. Fig. 5 is the schematic diagrams of the diagnostic models.

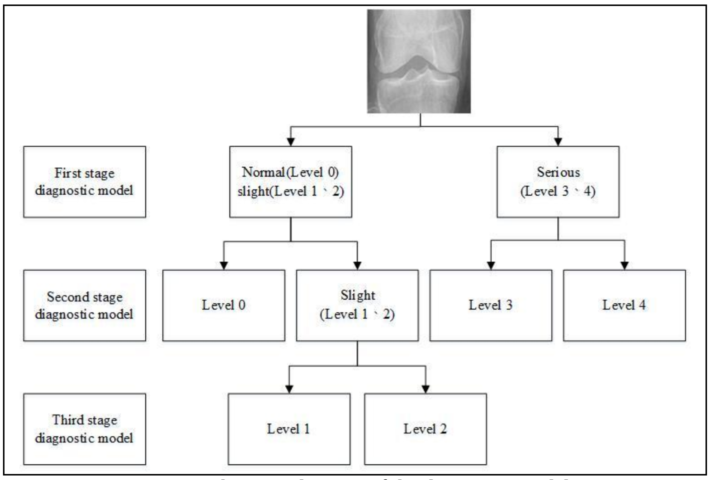

Fig. 5. Schematic diagram of the diagnostic model

1) The first stage of model training

In the first stage of model training, first, we classify K/L grade 0, 1, and 2 as slight categories and K/L grade 3 and 4 as severe categories for the training diagnostic model. Then, we can start training the first-stage model.

We use five pre-trained models to build the diagnostic model. After we train the models, we have to choose a suitable pre-trained model from them. We compare the accuracy and recall of five models using the same data set.

TABLE I is our experimental result in the first stage. They show the confusion matrix of the five models and a comparison of the accuracy of the five models. According to our experimental results, the recall and the accuracy of Inception V3 are better than other models, so we decide that Inception V3 becomes the model in the first stage.

TABLE I. Accuracy of five models for diagnosing in the first stage

S. No.	Model				
	Inception V3	**ResNet**	**DenseNet**	**Mobile Net**	**Efficient Net**
Accuracy	**0.9698**	0.9606	0.9658	0.9304	0.9448
Recall	**0.9237**	**0.9237**	0.9197	0.8835	0.8434
precision	0.9829	0.9426	0.9745	0.9016	**0.9859**

2) The second stage of model training

After the first stage model diagnosis, we need two models. The first model distinguishes normal and slight, and the other model distinguishes K/L grade 3 and K/L grade 4. Therefore, we will train these two models next.

The training steps are the same as the first stage. First, we also use five pre-trained models to build the diagnostic model. Then, we also compare the accuracy of five models using the same data set to choose a better pre-trained model from them.

TABLE II is our experimental results of the model to diagnose whether the knee joints are normal or slight. After Comparing the results, we can find that DenseNet has the best accuracy and the second-highest recall among the five models in the first model trained in the second stage. Therefore, we decide that DenseNet becomes the model that diagnoses whether osteoarthritis is normal or slight in the second stage.

TABLE II. Accuracy of three models for diagnosing in the first model trained in the second stage

S. No.	Model				
	Inception V3	**ResNet**	**DenseNet**	**Mobile Net**	**Efficient Net**
Accuracy	0.8301	0.8281	**0.8457**	0.7246	0.8066
Recall	0.7613	0.8230	0.8395	0.7778	**0.8519**
precision	**0.8645**	0.8163	0.8361	0.6848	0.7667

After training the diagnostic model that distinguishes between normal and slight, we began to train the diagnostic model which distinguishes between K/L grade 3 and K/L grade 4.

TABLE III is the evaluation of five models that diagnose whether the knee joints are K/L grade 3 or K/L grade 4. After comparing the results, Inception V3 has the best performance in the second model trained in the second stage. It has the best accuracy and the highest recall, so we decide that Inception

V3 becomes the model to diagnose whether osteoarthritis is K/L grade 3 or K/L grade 4.

TABLE III. Accuracy of five models for diagnosing in the second model trained in the second stage

S. No.	Model				
	Inception V3	**ResNet**	**DenseNet**	**Mobile Net**	**Efficient Net**
Accuracy	**0.9036**	0.8795	0.8956	0.8474	0.8313
Recall	**0.8817**	0.8172	0.8495	0.7957	0.7849
precision	0.8632	0.8539	**0.8681**	0.7957	0.7684

3) The second stage of model training

In the third stage of model training, we train a diagnostic model that can identify K/L grade 1 and K/L grade 2, after identifying slight categories and normal categories. We still train a model with five pre-trained models and then compare them to choose a better model to become the third stage of the diagnostic model.

TABLE IV shows the evaluation of five models. According to the comparison results, DenseNet has the best accuracy. Although there is no highest recall, we can find only a few deviations from the confusion matrix. Therefore, we decide that DenseNet becomes the diagnostic model in the third stage.

TABLE IV. Accuracy of five models for diagnosing in the third stage

S. No.	Model				
	Inception V3	**ResNet**	**DenseNet**	**Mobile Net**	**Efficient Net**
Accuracy	0.6626	0.6996	**0.7119**	0.6008	0.6049
Recall	0.7329	0.6957	0.7329	0.7764	**0.8012**
precision	0.7516	**0.8535**	0.8138	0.6720	0.6684

C. Interface designing phase

After the model is trained, the user interface of the diagnostic system must be designed. To let the diagnostic system be operated easily, it is designed that the user only needs to upload the knee X-ray images to the diagnostic system. Then, the system will diagnose the knee X-ray images, and present the results to the user.

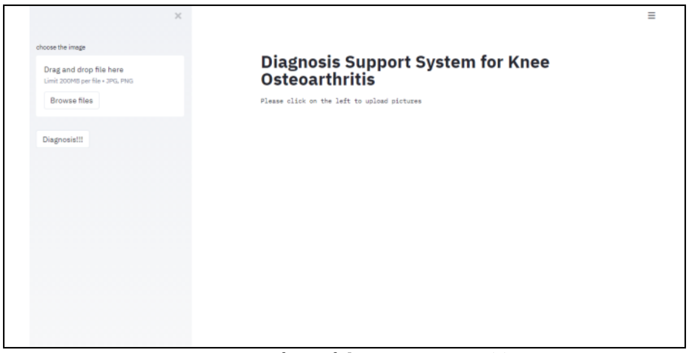

Fig. 6. Interface of diagnosis system(a)

Fig. 6 and Fig. 7, respectively, show the interface of the diagnosis system, which is divided into two parts, one is for selecting the knee X-ray image, and the other is used to diagnose the knee X-ray image and display the diagnosis result.

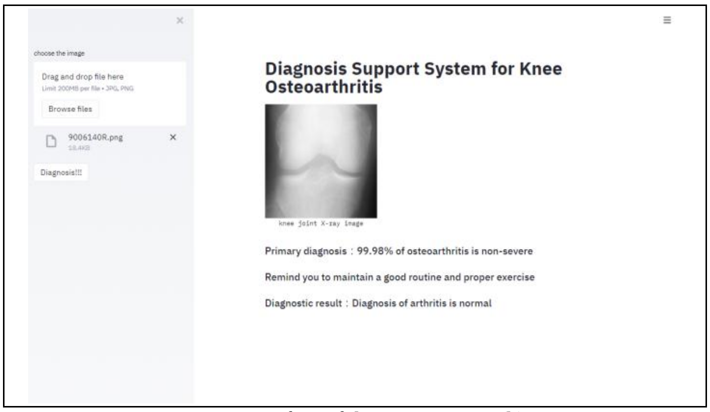

Fig. 7. Interface of diagnosis system (b)

V. Conclusion

This research shows that CNNs can be used to analyze the knee X-ray images. Moreover, the models proposed in recent years have performed well in this research. The diagnostic model at each stage of our system performs better than others.

This research focuses on distinguishing severe and non-severe knee osteoarthritis to assist doctors in filtering the less severe ones first, thereby reducing the burden on doctors and allowing severe ones to seek medical treatment first. The model in this study has the best performance at this stage. It can successfully distinguish joint conditions.

The ultimate goal of this research is to develop a medical support system. We can use our diagnostic system to assist doctors in analyzing knee X-ray images. The system can analyze knee X-ray images to know the severity of osteoarthritis in patients and recommend some suggestions for the doctor's reference.

References

[1] Hillary J. Braun, & Garry E. Gold, “Diagnosis of osteoarthritis: Imaging,” *Sciencedirect*, 2011.

[2] McAlindon T, &Dieppe P, “Osteoarthritis: definitions and criteria,” *Annals of the Rheumatic Diseases*, 48(7): 531-2, 1989.

[3] David Zelman, “How Do I Know if I Have Osteoarthritis,” *Webmd*, 2018.

[4] Kellgren JH, Lawrence JS., “Radiological assessment of osteo-arthrosis,” *Ann Rheum Dis*, 1957.

[5] Sharma, L., “Osteoarthritis year in review 2015: clinical,” *Osteoarthritis Research Society International*, 24(1), pp. 36-48, 2016

[6] M. C. Hochberg, R. D. Altman, K. D. Brandt, B. M. Clark, P. A. Dieppe, M. R. Griffin, R. W. Moskowitz, T. J. Schnitzer, “Guidelines for the medical management of osteoarthritis,” *Arthritis and Rheumatism*, 38(11), pp. 1541-1546, 1995.

[7] R. H. Osborne, R. Buchbinder, & I. N. Ackerman, “Can a disease-specific education program augment self-management skills and improve health-related quality of life in people with hip or knee osteoarthritis,” *Musculoskeletal Disorders*,7(90), pp. 1471-1487, 2006.

[8] Matthew A. Turk, Alex P. Pentland, “Eigenfaces for Recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[9] C. Farabet, C. Couprie, L. Najman, Y. LeCun, “Learning hierarchical features for scene labeling,” *IEEE transactions on pattern analysis and machine intelligence*, Vol. 35(8), pp. 1915-1929, 2012.

[10] J. J. Tompson, A. Jain, Y. LeCun, & C. Bregler, “Joint training of a convolutional network and a graphical model for human pose estimation,” *In Advances in neural information processing systems*, pp. 1799-1807, 2014.

[11] Li Deng, “Deep learning: from speech recognition to language and multimodal processing,” *Cambridge University Press*, 2016.

[12] Alessandro Palla, David Moloney, & Luca Fanucci, "Fully convolutional denoising autoencoder for 3D scene reconstruction from a single depth image," *2017 4th International Conference on Systems and Informatics (ICSAI)*, pp. 566-575, 2017.

[13] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, & L. D. Jackel, "Handwritten digit recognition with a back-propagation network," *In Advances in neural information processing systems*, pp. 396-404, 1990.

[14] Y. LeCun, L. Bottou, Y. Bengio, & P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, Vol. 86(11), pp. 2278-2324, 1998.

[15] V. Gulshan, L. Peng, M. Coram, Martin C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams & R. Kim, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, Vol. 316(22), pp. 2402-2410, 2016.

[16] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, Andrew Y. Ng, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," *arXiv preprint arXiv:*1711.05225, 2017.

[17] Joseph Antony, Kevin McGuinness, Noel E O Connor & Kieran Moran, "Quantifying Radiographic Knee Osteoarthritis Severity using Deep Convolutional Neural Networks," *arXiv preprint arXiv:*1609.02469, 2016.