Marquette University

e-Publications@Marquette

Master's Theses (2009 -)

Dissertations, Theses, and Professional Projects

Using Machine Learning Tools to Predict the Severity of Osteoarthritis Based on Knee X-Ray Data

YAORONG XIAO
Marquette University

Follow this and additional works at: https://epublications.marquette.edu/theses_open

Part of the Mathematics Commons

Recommended Citation

XIAO, YAORONG, "Using Machine Learning Tools to Predict the Severity of Osteoarthritis Based on Knee X-Ray Data" (2020). *Master's Theses (2009 -)*. 582.
https://epublications.marquette.edu/theses_open/582

USING MACHINE LEARNING TOOLS TO PREDICT THE SEVERITY OF OSTEOARTHRITIS BASED ON KNEE X-RAY DATA

By

Yaorong Xiao

A Thesis submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin
May 2020

ABSTRACT
USING MACHINE LEARNING TOOLS TO PREDICT THE SEVERITY OF OSTEOARTHRITIS
BASED ON KNEE X-RAY DATA

Yaorong Xiao

Marquette University, 2020

Knee osteoarthritis(OA) is a very general joint disease that disturb many people especially people over 60. The severity of pain caused by knee OA is the most important portent to disable. Until now, the bad impact of osteoarthritis on health care and public health systems is still increasing.

In this paper, we will build a machine learning model to detect the edge of the knee based on the X-ray image and predict the severity of OA. We use a clustering algorithm and machine learning tools to predict the severity of OA in knee X-ray images. The data is coming from the OsteoArthritis Initiative (OAI). To process the data, we use the clustering method as the first step to do unsupervised learning on the dataset and get clusters from each single X-ray image. For every single image, we can get features. Therefore, we transfer complicate image data into simple data, a vector. Then, we use machine learning tools to analyze the extracted feature data and detect the severity of knee OA. We also built a convolutional neural network (CNN) model to make a comparison between the method we used and deep learning algorithm.

Key words: machine learning, X-ray, Osteoarthritis

ACKNOWLEDGMENTS

Yaorong Xiao, Naveen Bansal

TABLE OF CONTENTS"

LIST OF TABLES

LIST OF FIGURES

1 Introduction

Knee osteoarthritis (OA) is a highly prevalent condition that will gradually cause

disability, which troubled older adults a lot (Medina., 2016). Knee OA is a complicated disease to

cure and hard to detect at the beginning of the disease. The prevalence rate of knee OA is very

high, which occurs in 10% men and 13% in women aged 60 years or older (Zhang and Jordan,

2010). Until now, nuclear magnetic resonance (MRI) is the most useful method to detect knee OA

(Oo, et al., 2017). There are many proofs that the pain of knee joint has a significant correlation

with the MRI features (Javaid, et al., 2012; Minciullo et al., 2018). Generally, only more

professional doctors with rich experience can detect knee OA with a patient's periodical reaction

and X-ray image. That wastes a lot of time and workforce. Also, it will increase the cost. What is

more, many researches show that the severity of knee OA related to their individual beliefs

(Darlow, et al., 2018). So, finding a way to detect knee OA is a necessary problem.

With so many shortcomings, in recent years, many scientists participated in OA detection

work. Many machine learning and deep learning models have been used to detect knee OA, such

as building a convolutional neural network (Antony et al., 2017) or presenting a computer-aided

diagnosis system (Brahim et al., 2019). Building deep and complicated CNN is the most popular

method to make the prediction. However, the training process requires much data, and sometimes

it will take several months to train. Therefore, in this paper, we focus on using machine learning

tools except building a complicated neural network. However, we still face many difficulties. The

different brightness of X-ray images causes a similar severity to have utterly different pixel

values. The overlap between bones will affect the accuracy of clustering. Furthermore, the

dispersive noise point will also influence the classifying process.

Before, people estimate the displacements from image pitches via a voting scheme to make landmark detection and shape segmentation (Chen et al., 2014). While in this paper, we use an unsupervised clustering method to detect the joint displacement. Given an X-ray image of Knee, the most intuitive detection of OA is to calculate the distance of the bones between edge. To do this, we cut the X-ray image vertically, divide the single image into two pictures of the same size, and do cluster analysis on both of them. Then images with same severity will have similar cluster. The features are extracted from these two separated images and will be used for classification to the next step.

We used multiple methods to detect the severity according to the extracted features. Logistic regression is the first and most straightforward method we used. We also used naive Bayes classifier and k-nearest neighbors algorithm (k-NN) and finally chose the k-NN for model use. K-NN algorithm is a fast and straightforward non-parametric method to classify the severity of knee OA and only takes a few minutes to train. Compared with CNN, which will take more than ten hours to train, k-NN saves more time and finally get a similar result. We built an Xception network Chollet, 2017), a deep and complicated convolutional neural network for comparison. What is more, in order to improve the time to run CNN, we use Google Colab as the environment. Google Colab is a browser that allows people to write and execute Python online.

In this paper we also met some problem for image processing. Noise in the x-ray images is one of the most serious one. In the process of image acquisition, transmission and storage, it is often interfered and affected by various noises to degrade the image, and the quality of the image preprocessing algorithm is directly related to the effect of subsequent image processing, such as

image segmentation, Target recognition, edge extraction, etc., in order to obtain high-quality

digital images, it is necessary to reduce noise on the image, as far as possible to maintain the

integrity of the original information while removing unwanted information from the signal. For

the knee OA data, noises are randomly located on the image. And we will show what happened in

the discussion part.

 The paper structure is: Firstly, we briefly describe the related work we have done.

Secondly, introduce the CNN we used for comparison. Then we go through the method we made

in this paper. Next, we will present the experiments. The final part is the conclusion and

application.

2 Related work

Since the knee-joint has the largest joint surface and bears most of the body's weight, there is no doubt that the knee-joint is one of the most vulnerable joints in the human body (Henche, 1980). In America, there are nearly 2 million primary care outpatient visits are due to acute knee pain (Jackson et al., 2003). At the early time, MRI is the most popular method for experimental doctors to detect knee OA. However, the decision of the final diagnosis is based on high cost and have lower accuracy. To accurately detect the knee OA severity is necessary for patients. A better understanding of the disease can lead to timely prevention and treatment of knee OA, one of the most common causes of disability in adults.

Some methods have been used to analyze X-ray images of the knee at early stages to predict whether it would have OA in the future. The authors use a computer-aided image method to predict the subtle difference in textures and intensity variations within an image, without clinical bias (Shamir et al., 2009). This method performed well. It has a 72% and 62% accuracy on detecting whether the severity would change from grade 0 to grade 3 and grade 2 respectively. Generally, the X-ray image of grade 0, 1, and 2 are very similar. 72% accuracy is an excellent result of this prediction. They firstly use a WND-CHARM algorithm (Shamir et al., 2008) to extract image features, which provide a numeric description of the image.

Detecting knee OA at an early age is very necessary for curing and prevention. Therefore, a more accurate prediction is needed. With the advent of the era of big data and the significant development of technology, Machine learning becomes very popular—many machine learning tools, such as decision tree, random forest, linear regression and multilayer perceptron neural

network (Aljaaf et al., 2016), used in industries and hospitals. So many scientists started to try using these methods to detect the severity of knee OA. Mostly, scientists detect knee OA by using deep learning methods. A deep and complicated convolutional neural network can have a high performance in X-ray image analysis, especially in disease detection. For instance, people built a novel end-to-end CNN architecture to automatically detect the severity of knee OA (Gorriz et al., 2019; Lim et al., 2019). They get a 64.3% TEST accuracy by using VGG-16 (Simonyan et al., 2015). Furthermore, CNN also used in analyzing many other X-ray images such as pneumonia detection (Tsai et al., 2019). While the problem is that people still not know how this neural network does in the 'black box' (Bleicher, 2017). So, we are looking for other methods instead of using CNN. In recent years, there have been many people using machine learning tools to predict knee OA severity. Landmark detection and shape segmentation are some of the most popular methods used in this problem (Tiulpin et al., 2019). This method detects the landmark defined along to the shape contour first and gets the regularized landmark detection output. Then use both the output and prior shape information to get the final contour (Linder et al., 2013). While what we have done here is that we made the problem becomes more straightforward than before. Abandon the complicated and massive calculation of landmark detection. We use the clustering method and acquire the pixel value at the joint site. We did not detect the landmark. Instead, we use the clustering method to generate a cluster that has the boundary as the 'landmark.'

CNN is often the final choice for most people who are working to detect knee OA. And many process algorithms, or machine tools, are used to process the original data. Then the processed data will finally send into a neural network. Only recent work abandoned it and replaced the CNN with Naive Bayes (Du, 2018). Except for using the deep learning method,

many other machine learning tools also used in the detection of knee OA. For example, a decision

support tool is used as the initial detection of the knee X-ray image, and then the machine

learning tool will be used for classification (Brahim et al., 2019). The first step of this method is

to use a circular Fourier filter. Then using independent component analysis (ICA) to extract

features and finally, Naive Bayes and random forest classifiers are used for the classification task.

The article has an accuracy of 82%, which is an outstanding performance. In another literature, it

used a support vector machine (SVM) to compare with the linear discriminant analysis (LDA)

(Moustakidis et al., 2019). This literature also tests many other machine learning algorithms and

deep learning tools. What we do in this paper is different. We use clustering method to pre-

process the data and instead of using circular Fourier filter we use machine tools to classify. In the

beginning, we plan to use a multidimensional Gaussian distribution to process the X-ray image

(Li et al., 2016). However, problems exist in the many original images. Because the brightness of

images depends on the value of each pixel, the brightness of X-ray images is different for each

individual. So, doing multidimensional Gaussian distribution on images with the same severity

will have a different result. To overcome this problem, wo choose to use clustering method.

Clusters have more specific and clearer categories. The features extracted are more significant.

3 Data

3.1 Data Resource

All the X-ray knee data are downloaded from The Osteoarthritis Initiative (OAI). The OAI is a nationwide research study, sponsored by the National Institutes of Health (https://www.nih.gov/) (2018, January 5th). This data set contains both knee joint detection and Kellgren-Lawrence (KL) grading (Kohn et al., 2016). And it also has 224 times 224 (number of pixels) and 299 times 299 versions of the image data. What we did in this paper is to predict the grading level by using machine learning tools.

3.2 Normalization

In this part, we built a deep and complicated convolutional neural network. Before sending the image data to the network. We need to normalize the value of all the pixels due to the brightness problem we mentioned before. There are some different normalized methods, and what we use in this paper is:

$$y_i = \frac{x_i - \bar{x}}{s}, here\ \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\ , s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1}$$

In this function, $\bar{x}$ represents the mean value and s represents the standard deviation. x_i represents each single-pixel value with i∈[1:224*224]. While y_i is the normalized new value used for training. This normalized method is based on the mean and standard deviation of the raw data. It uses the z-score to normalize the original data to new data.

4 CNN method for OA detection

Xception is an upgraded version of Inception V3 (Chollet, 2017). This new CNN has the same number of parameters as the Inception V3 but does have more effective use of these parameters. Xception replaces the convolution part in Inception V3 into using a depthwise separable convolution. This convolution can reduce the number of parameters, meanwhile increase channels and feature expression. We first go through the Inception V3:

Figure 1
A Canonical Inception Module (Inception V3)
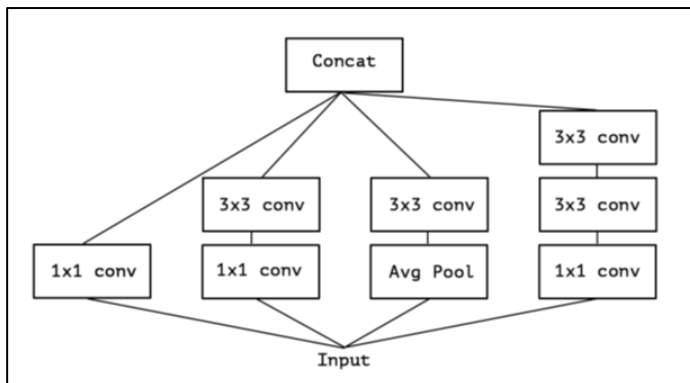
Note: A canonical Inception module (Inception V3). Adapted from 'Xception: Deep Learning with Depthwise Separable Convolutions' by Francois Chollet, 2017.

The basic idea of Inception V3 is: input the data into different convolution structures at the same time to extract features and then concat. The Inception V3 changes the 5*5 kernel into two 3*3 kernels. This neural network can be simplified. In figure 2, using a 1*1 kernel as the first layer and then connect to 3*3 kernels. Let's consider an extreme condition. When the number of 3*3 convolution equals to the output channel of the 1*1 layer, we can get:

Figure 2
A Strictly Equivalent Reformulation of the Simplified Inception Module.
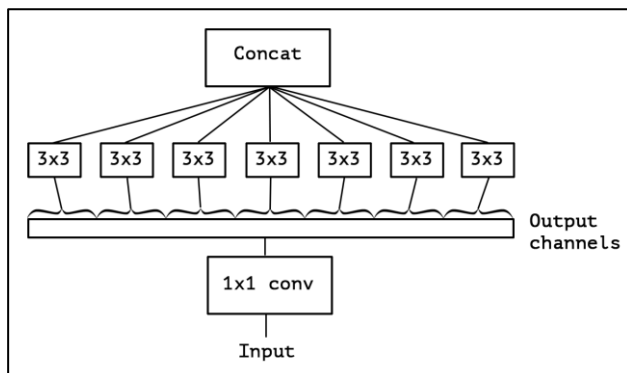
Note: An "extreme" version of our Inception module, with one spatial convolution per output channel of the 1x1 convolution.

Adapted from 'Xception: Deep Learning with Depthwise Separable Convolutions' by Francois Chollet, 2017.

The primary process of depthwise separable convolution is: divide the traditional convolution neural network into two steps. Assume the original kernel is 3*3. Then for depthwise separable convolution, it will first use M (the number of input channels) 3*3 kernels to filter each input channel. It gets M output. Next, using N (the number of input) 1*1 kernels to filter the M outputs and get N outputs. These two steps are called depthwise convolution and pointwise convolution, respectively. Thus, the structure of the Xception is in figure 3. The sparsableConv is the depthwise separable convolution. The mark '+' means residual connection

Figure 3.
The Xception Architecture

Entry flow   Middle flow   Exit flow

Note: the data first goes through the entry flow, then through the middle flow which is repeated eight times, and finally through the exit flow. Note that all Convolution and Separable Convolution layers are followed by batch normalization (Ioffe et al., 2015) (not included in the diagram). All Separable Convolution layers use a depth multiplier of 1 (no depth expansion). Adapted from 'Xception: Deep Learning with Depthwise Separable Convolutions' by Francois Chollet, 2017.

The Xception is a very deep and complicated CNN with 36 convolutional layers. The input should be 299*299 images. To run it fully, we need a laptop with a great GPU. And finally, we chose to use Google Colab, a browser that can run Python code. It has a powerful graphics processing unit (GPU) and significantly improves the program speed. The data set are zipped in a zip file and uploads to the google drive. From the output in table 1, we can see that the accuracy of training data convergence at around 66%. This neural network takes 10 hours to run. And the accuracy of test data is 0.626.

Table 1

Error and Accuracy of Xception

Epoch	Error	Accuracy of training data
1	1.428989	0.40714
2	1.332603	0.39429
3	1.544461	0.39286
4	1.387053	0.41571
… …		
57	1.210080	0.67857
58	1.339737	0.64000
59	1.284551	0.64571
60	1.105308	0.65714
61	1.155554	0.63571
62	1.210734	0.66429
63	1.256519	0.66286
64	1.202377	0.65286
65	1.245270	0.67000

5 K-MEANS CLUSTERING METHOD FOR OA DETECTION

In this study, we use the K-means clustering algorithm to process the image data first. Some people have used a clustering method to analyze the dental X-ray image to obtain the teeth contour (Rad et al., 2013). Being illuminated by this article, it is also possible to use the clustering method to analyze the knee X-ray image and extract the joint space features. Similar to the Xception, in the clustering method, we also do normalization for each image data due to the brightness problem. In each image, different areas, joint space and bones, have different size unit. In this condition, normalization can be used to reduce the effect of data analysis, eliminate dimensional effects between index   es, and improve comparability between indexes. It is obvious that in figure 4, the left image is light with higher pixel intensity values. The right image is darker with lower pixel intensity values.

In (1), we have showed what normalization method we used in this paper. Let's see what happened to the normalized data before and after. We can see that, there is no significant different between the two images in figure 5. For the y-axis, the value has changed from a range of 0 to 255 to very small range number. While the trend of the pixel values does not change. In this figure, there are 50176 pixels, and from approximately number 10000 to number 30000. The pixels are counted by columns. That means that the first 224 data are counted from the first column of the image then the second column. In view of this, it is clearly shown in the figure that at the start and the end, the value is relatively low. And at the center of each images, there is a slight inflection where is the joint space exists. So, our goal is to detect this part by clustering method.

Figure 4

Brightness Comparison of Two Images
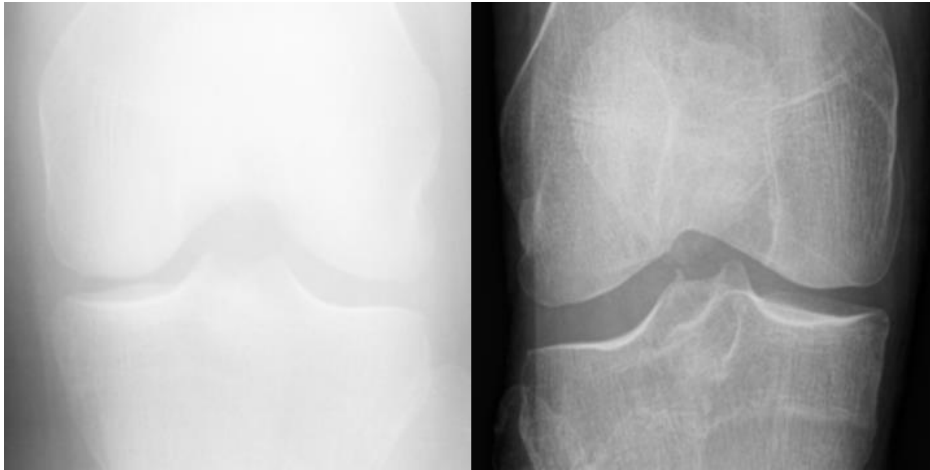
Note. These two pictures are randomly picked from the OAI data set.

Function 2

K-means Algorithm

Step 1. Randomly pick k cluster centroids $\mu_1, \mu_2, \dots \dots, \mu_k$

Step 2. Repeat the following process until convergence:

For each sample i, calculate what category it belongs to:

$$C^{(i)} := arg \min_j \left\| x^{(i)} - \mu_j \right\|^2$$

For each category j, recalculate the cluster centroid:

$$\mu_j := \frac{\sum_{i=1}^{m} 1\{c^{(i)}=j\}x^{(i)}}{\sum_{i=1}^{m} 1\{c^{(i)}=j\}}$$

Different to regression, naïve Bayes and support vector machine (SVM), K-means clustering is unsupervised learning used to draw inferences from datasets consisting of input data without labeled responses. Because the object we used in the clustering method are the pixel values. So, we do not need any label of the pixels by using this method. The purpose of clustering is to find the potential category of each given data and put it together. For K-means clustering method to detect knee OA. The training sample we have is $\{x_1, x_2, x_3, x_4, \dots, x_{224*224}\}$. K-means

clustering requires to classify the 50176 pixels into pre-defined clusters. From figure 5 we can see

that the brightness of the joint space is significant different to the upper and below bones. X-ray

images are one-dimensional image. The darker the image is the smaller pixel value it has.

Figure 5
Scatter Plot for Normalized and Nonnormalized Data

Note. These two images are generated by R.
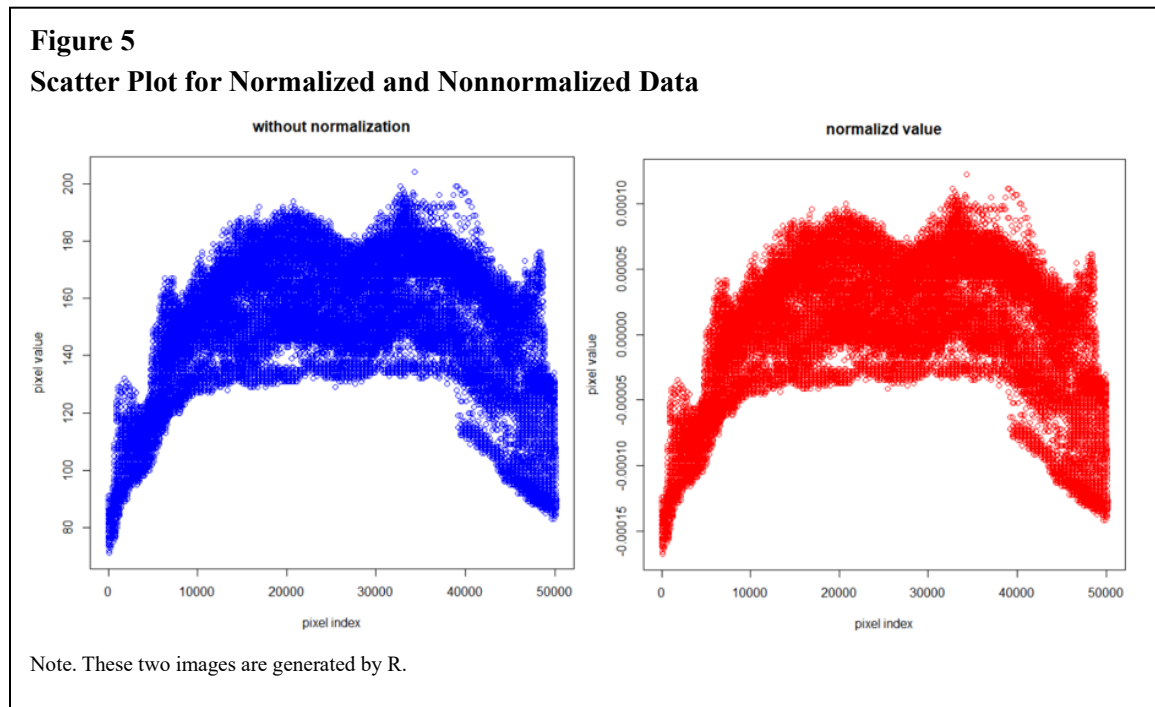
The algorithm of K-means shown in function 2. k is the number of clusters we preset,

$c^{(i)}$ represents the nearest cluster to sample i. $c^{(i)}$ is an integer belongs to 1 to k. Then each of the

sample data belongs to one cluster. Using the classified data, recalculate the new cluster centroids to

replace the original one. Repeat this process until μ_k convergence.

We know how the K-means clustering works. One of the essential things for processing

data is to decide the number of cluster centroids. By using clustering method, we hope the pixel

values of joint space be strictly clustered. So, selecting a unique centroid of the joint space is the

key to this problem. For image 1 (right), because of the weak difference of the edge pixel values,

he joint space sometimes will be mixed with some part of the bones. Therefore, we can see that

two clusters would be the best choice. Because joint space, bones, and the unrelated background

have apparent different brightness. However, for the left one, the brightness of joint space and

background are nearly the same. So, if we want to have an accurate cluster of the joint space,

removing the background pixels is the best choice. Going through the whole 9786 images, the

size of the knees in the images are the same.

Figure 6
Comparison Between Different Number of Clusters and Clipped Image

Note. The left image is the original clustered image. The middle one is clipped image. The right image is the result by using 3 clusters.

So, we can manually cut the image to get a new one. Before manual work, we tried to use

edge detection to cut the image. While due to the nuance different among images. Using edge

detection cannot guarantee the same strict size for each image. Because we want to extract

features from each image, the same size is necessary for the final comparison. While we do obtain

some useful information from edge detection, most of the images have around 154 columns. So,

we decide to cut the image from column 35 to 189, and the size of the X-ray images becomes

224*154. Then, we can do a K-means clustering with 2 centroids. The left and middle images of figure 6 show the difference between the original image and the cut image. If we use three clusters, there is a larger error to the joint space cluster. Because the pixels at the edge of bones assigned to that cluster, see figure 6 (right), we choose to use 2 clusters for data pre-processing.

Figure 7
Folded Images in Different Condition

Note. These three images are generated by R.
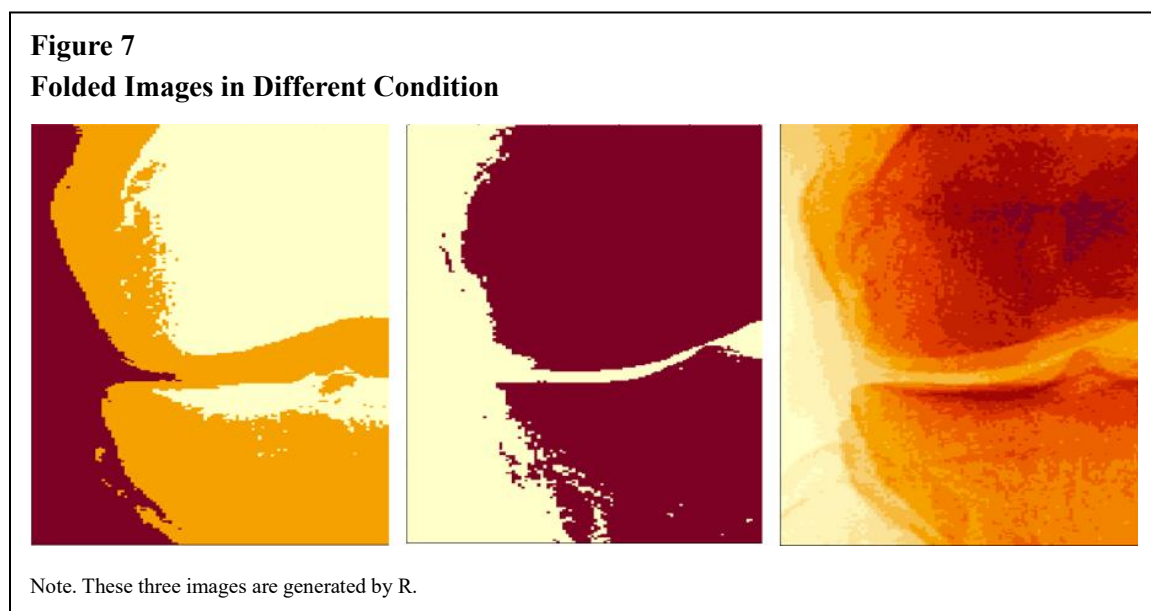
What is more, we also tried other method to eliminate the background of images. What we do is folding the images left and right. As we know, most people's knee is highly symmetrical. So, the folding process will to some extend increase the contrast between joint space and tissue bones. What we except is the joint part can be clearly divided. However, according to figure 7 (left), because of the folding process, the part of the edge of bones also become significant with 3 clusters. Then we tried 2 clusters, the result is the middle one of figure 7. In terms of the left and middle images, we found for two clusters, some part of bones is clustered into the joint space cluster. The error is significant increased. Furthermore, we can see through the right image that it

is no strictly equally divided. This could be caused by the knee position when taking the x-ray or

the skeleton frame. And the scatter plot is also shown below. Compared with figure 5, the trend of

them are nearly same. While we can see that at the end of folded scatter plot. The joint space is

not distinct. So, due to the many problem that could impact our final result. We decided to

abandon this method and use another cutting method to extract features.

Figure 8
Scatter Plot of Folded Image

Note. This image is generated by R
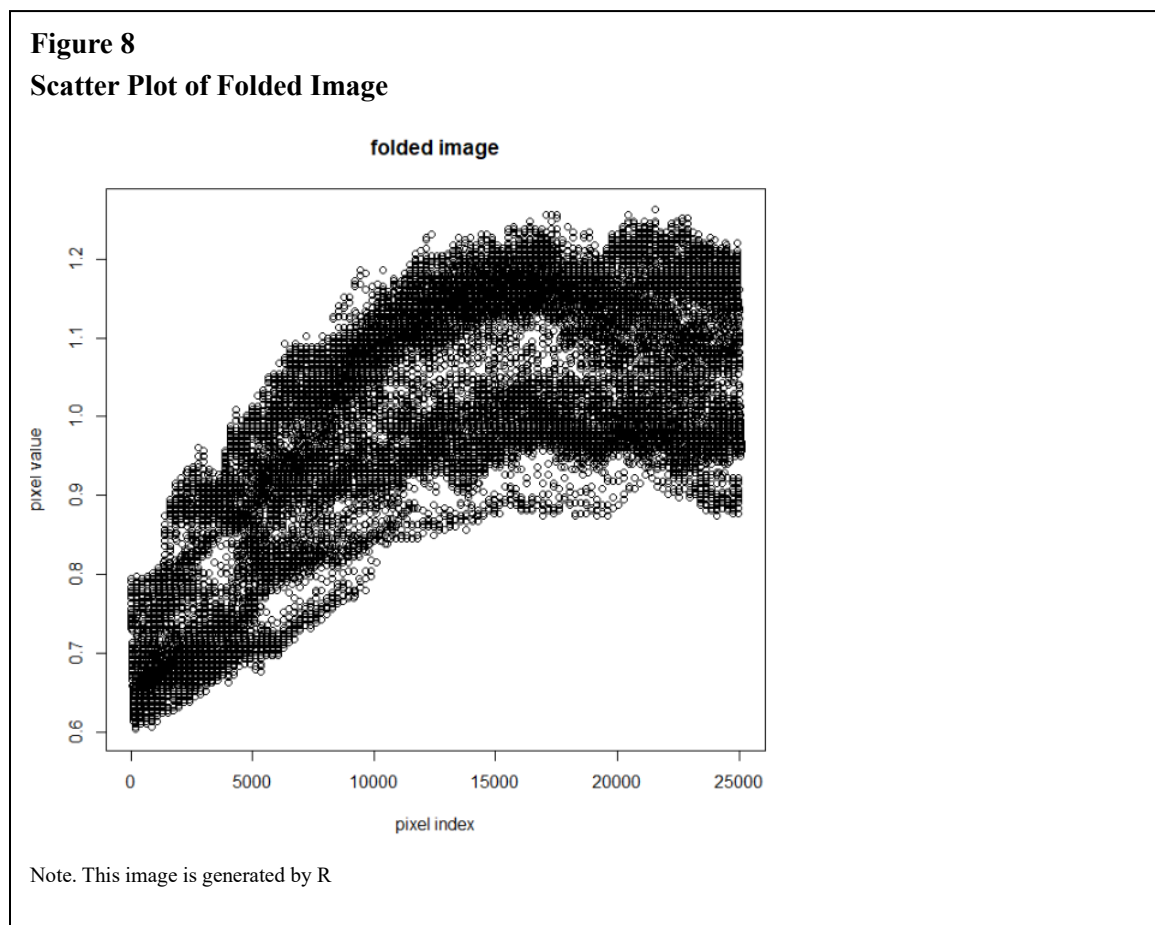
As we have done, we can obtain two clusters from one image. However, in this condition,

we cannot extract enough significant features. As we know, All the X-ray images are collected

from one single person. Which means the data are mixed with left and right legs. So, for every

single image (no matter left or right leg), the left part of it has a slight difference compared to the

other side of this image. Thus, what we can do here is to divide the image into left and right parts

equally. Then doing K-means clustering on both of them. The most general method to detect knee

OA is to detect the area of the joint space. We have already obtained the mean value of the joint

space from the clustering method. And we also have the number of pixels in each cluster. Thus,

we can calculate the area of joint space. Because of the overlap of joint space and bones, we also

calculate the standard deviation the cluster of the joint space due to the obscure pixel value of the

edge point between joint space and bones.

Back to how to detect the severity of knee OA, the larger space the joint has, the higher

the OA level it is. So, what we want to do now is to calculate the summation of the pixels at the

joint space. The number of pixels and the mean of these values have already obtained from the

clustering method. Then, for each original image data, we can obtain 5 features: the mean and the

standard deviation of the joint space of both left and right sides. We also append the difference

between the two mean values as the fifth feature. So, we obtain a five-dimensional vector from

each image. We use these feature vectors as the source for classification.

The classification method we used in this paper is the k-nearest neighbors algorithm

(KNN). We already get a data set of five-dimensional vectors from the original data. The

definition of KNN is: In a feature space, if the k nearest points of the input belong to category X.

Then the input belongs to X. This is a non-parametric method. Because the 5 features do not have

significant weight, so we choose the non-parametric classification method. KNN is not sensitive

to the outliers, and in our data set, there are abnormal image exists. So, we use KNN to do

classification. The method we used in this paper is a binary classification algorithm. The knee OA

have five severity levels, represented by numbers from 0 to 4. 0 means normal knee and 4 means

the worst OA condition. We do binary classification by comparing each level to the corresponding

more serious level. For instance, defining 0 as a group and the remaining data as another group.

What it means here, we compare level 0 with level 1 to 4, and then compare level 1 with level 2

to 3. For this problem, we the more serious OA as the compared data set. If we use level 1 as one

category, level 0 and level 2, 3, 4 as another category. The feature set of the last category is not

significant since level 0 is more similar to level 1 rather to the other levels.

We also tried different classification method, such as support vector machine (SVM), The

definition of SVM is: The objective of the support vector machine algorithm is to find a

hyperplane in an N-dimensional space (N-the number of features) that distinctly classifies the

data points (Gandhi, 2018). SVM is a novel small sample learning method with a solid

theoretical foundation. It basically does not involve the probability measurement and the law of

large numbers, etc., so it is different from the existing statistical methods. Essentially, it avoids

the traditional process from induction to deduction, and achieves efficient "transduction

reasoning" from training samples to forecasting samples, which greatly simplifies the usual

classification and regression problems. What we have in this paper, we transfer each single image

(with 50176 pixels values) to a 5-dimensional vector. SVM adopts the hinge loss as its loss

function, the purpose of this loss function is to increase the weight of the features which has a

larger effect to classification. The basic idea of SVM is to find the hyperplane. However, for

linear separable data, such kind of hyperplane have a lot. But the larges geometric distance

hyperplane is unique. According to figure, the difference between level 0 and level 4 can be seen.

So, it is possible to use SVM to classify the severity of it.

Our data are five-dimensional data, it is not easy to exhibit the data as image. We use python and generate an SVM model. The results are shown below. From the table, no matter what comparison listed below. We can see that the largest accuracy is 0.52. This means this method does not have effective result. For this process, normalization is also used due to the brightness of images. From the theory, it is possible that we can build a more specific hyperplane to divide data. However, if it does not work in this condition. There should be some reason exist in the data. We also tried another normalization method. That is stressing all the data proportionally between 0 to 1. While the scatter plot is still the same compared with another normalization we used before.

Table 2

Results of Different Comparison of Severity by SVM

Experiment	Data set for comparison		Accuracy
1	Level 0	Level 1, 2, 3, 4	0.4715
2	Level 1	Level 2, 3, 4	0.5281
3	Level 2	Level 3, 4	0.4216
4	Level 0	Level 3, 4	0.3056
5	Level 0	Level 2, 3, 4	0.4676
6	Level 0, 1	Level 2, 3, 4	0.4815
7	Level 0, 1	Level 3, 4	0.3546

6 Discussion

We performed two main methods in this paper to detect the severity of knee OA. Cluster analysis is the method we want to post in this paper. The results are showing in table 3. As we introduced before in this paper, we use binary classification methods, training, and testing the data set by comparing each level to the corresponding more severe level. According to table 3, we can see that the highest accuracy is the comparison between the level 0 to level 3 and 4. It shows a 67.25% accuracy. Firstly, we classify the image by detecting whether the knee has OA. So, we divide the data set into level 0 and level 1 to 4 to do classification. The result we obtain in this condition is 55.37%. That is not a satisfying result. So, we change to detect level 1 to level 2, 3, 4. Then, we can observe that there is a relative improvement on the result with an accuracy of 61.16%. After a few times try, we realized that level 0 and level 1 have a similar look. It is hard work to discriminate the difference between them. Thus, we merge level 0 and level 1 into one group and the remaining 3 levels as another group. The result we get shown in table 3, and the accuracy is 56.15%. We can find that the results of experiment 1 and experiment 6 are the same. Also, in experiment 5, the accuracy is even worth with an accuracy of 54.12%. In this condition, we can see that the data set of level 2 is weird. When experiments contain level 2 for classification, the result is not accurate. By looking at the whole data set, the significance of level 2 is, to some extent, similar to the image of both more and less severe images. Therefore, we decide to get rid of level 2 for detection. In experiment 4, we use group 1 (include level 0) and group 2 (include level 3 and 4). The result has a significant improvement with 67.25% accuracy. On this basis, we did experiment 7. With level 0 and 1 as a group, level 3 and 4 as another group,

finally, we obtain 72.64% accuracy.

Table 3
Results of Different Comparison of Severity by KNN

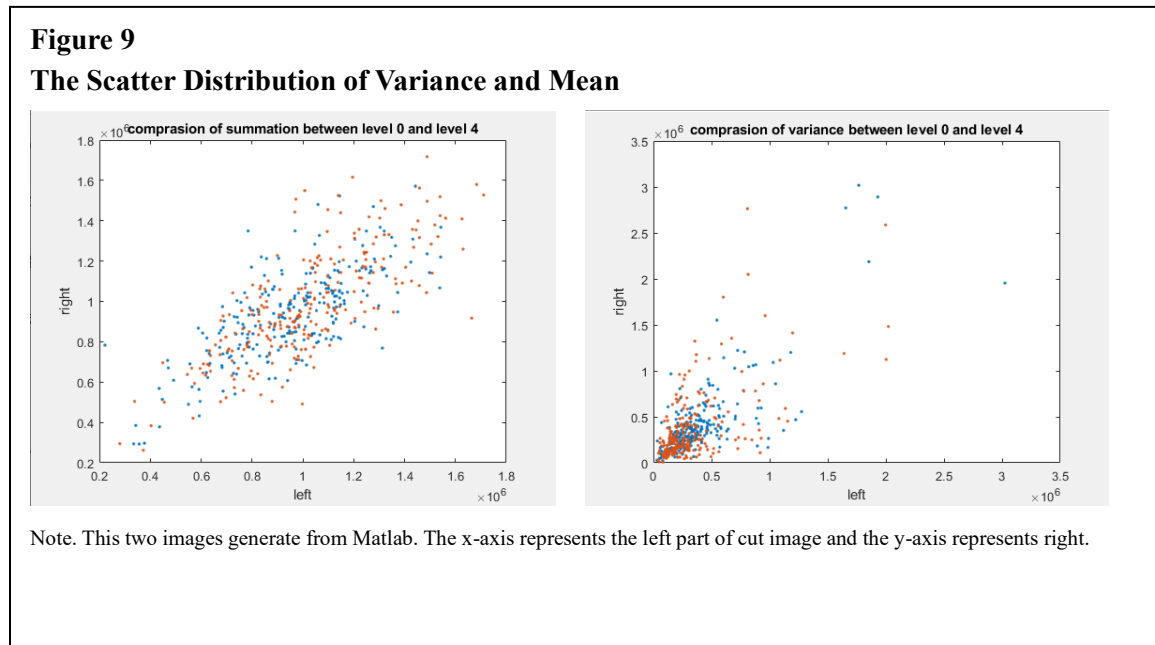| Experiment	Data set for comparison		Accuracy
1	Level 0	Level 1, 2, 3, 4	0.5537
2	Level 1	Level 2, 3, 4	0.6116
3	Level 2	Level 3, 4	0.6227
4	Level 0	Level 3, 4	0.6725
5	Level 0	Level 2, 3, 4	0.5412
6	Level 0, 1	Level 2, 3, 4	0.5615
7	Level 0, 1	Level 3, 4	0.7246

Table 4
Mean Value of the Features of Level 0 and Level 4

	Summation of left	Summation of right	Standard deviation of left	Standard deviation of right
Level 0	0.9544	0.9327	0.3973	0.4184
Level 4	1.0156	0.9796	0.3380	0.3767

Note. All the value in this table are represent by scientific notation with accuracy of 10^6.

The figure 9 is the scatter plot of mean and variance of level 0 and level 4. According to the figures, although we see that the cluster mean of these two levels are not significantly different. While according to table 4, we can see that the value of table 4 do have a significant different since it uses scientific notation. Looking back to the summation figure again, we can see that there do have a slight difference of the two cluster. The red cluster is more up and right. On

the other hand, for the variance value, the differences are clear. These two figures are chosen from

level 0 and level 4 because the difference is more distinct.

Figure 9
The Scatter Distribution of Variance and Mean

Note. This two images generate from Matlab. The x-axis represents the left part of cut image and the y-axis represents right.

6.1 Reason for Using K-Means Clustering Algorithm

Building a neural network is mostly used to resolve high-dimensional data. High-

dimensional data can be converted to low-dimensional codes by training a multilayer neural

network with a small central layer to reconstruct high-dimensional input vectors (Hinton, 2006).

In that article, it proves that as long as the layer of a neural network is deep enough, the neural

network will have higher efficiency in extracting features than PCA. It also means that, as long as

we have sufficient computing resources, we even do not need to use the statistical method to

analyze data and find features. Compared with the traditional machine learning algorithm, deep

learning algorithms focus on extracting features from massive data and resolve the high-

dimensional, verbose, and high-noise problems existing in massive data which are difficult to be

dealt with by traditional machine learning algorithms. Deep learning is to design a machine

learning algorithm to automatically learn the knowledge representation of data and the

complicated relationship between knowledge representations from massive data. Based on these

knowledge representations and relationships of complete pattern recognition tasks, the ultimate

goal is to make the machine achieve real artificial intelligence.

Figure 10
Black Box Exhibition
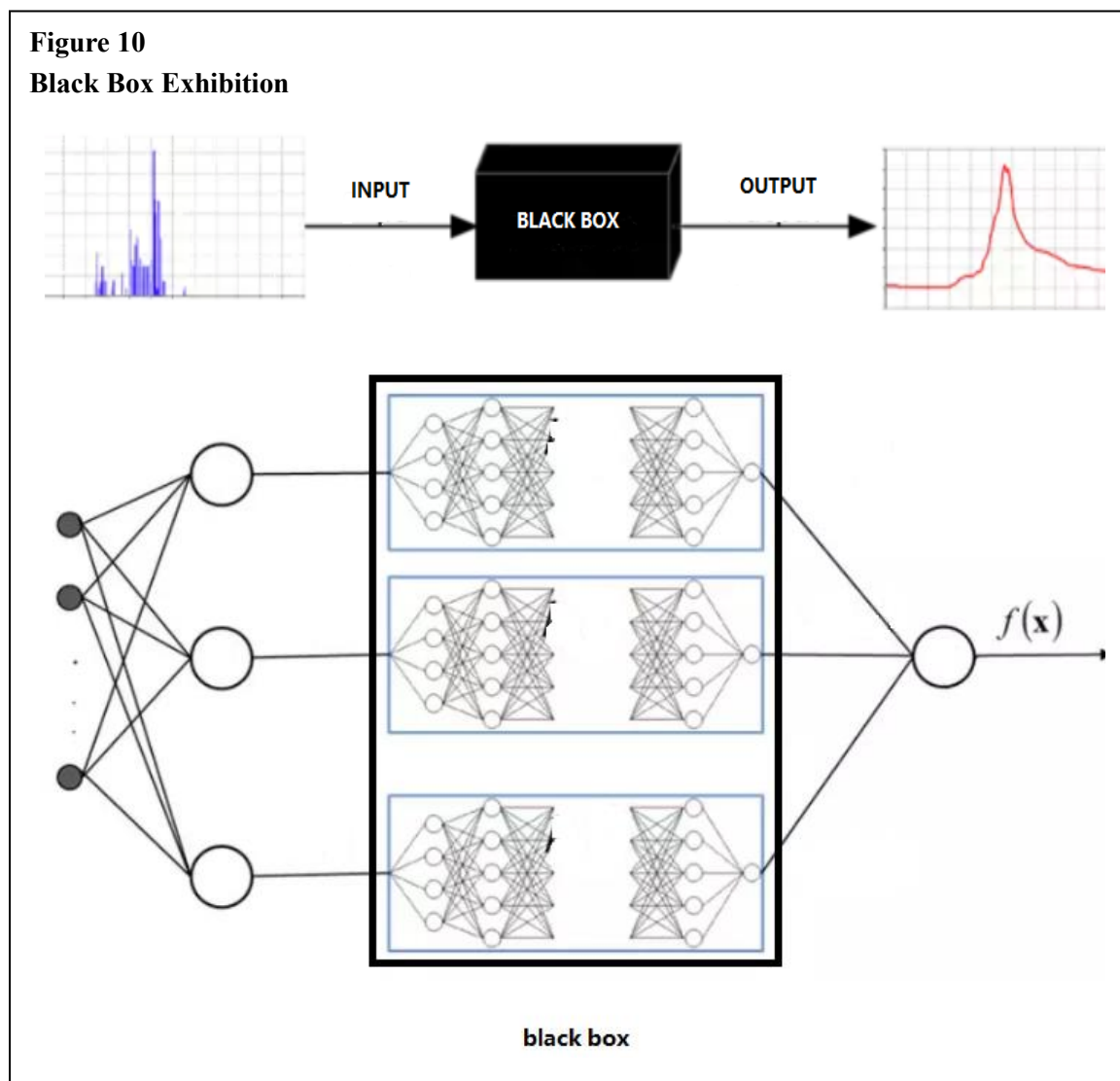
However, it is true that for a human himself, he does not know how exactly his brain operates during the thinking and working process. That leads to a problem of deep learning, more precisely, a convolutional neural network. The black box problem always exists in a deep learning algorithm. We know that deep learning algorithms can help us to solve many big data problems, such as image analysis and natural language process. However, how it works is still a mystery. We know that CNN uses many statistical tools to improve the coefficients to reduce the error. Whereas for a complicated and deep neural network, especially for convolutional neural networks, there are millions of coefficients, sometimes more, exist. With the number of layers increase, the neural network will also become more complex and massive. People cannot detect the specific correlation between each node and layers; what they know is that this method could work and have a high performance. It can be seen that Neural networks cannot learn knowledge directly. Features are extracted from data. If people want to teach human knowledge directly, such as Einstein's theory of relativity, there is no corresponding solution. Therefore, in our paper, we used other machine learning method.

Many people and literature have researched on the knee OA data. The experience of this literature and doctors, especially, tell us that the area of joint space reflects the severity of knee OA (Lewis et al., 2012). CNN did an excellent job on OA detection. However, we never know how it works on the x-ray images. All the images are entered as input and exited as a detailed result. We never what happened on each pixel value of the image data. We do not know how to define a coefficient at a specific layer's specific node. It seems that the results are the most important thing for the OA detection process. However, the algorithm and the detection process will be used on medication care in the future. So just a single result will not have any analysis

effect for patients. As mentioned in this paper before, the general method to detect the severity of

knee OA is to detect the space of the knee joint. Since we do not know what CNN did in the black

box during the training process, the features that CNN extracted may be more than joint space. In

terms of this condition, we decide to find an algorithm to find the area of joint space.

According to many works of literature listed before, people have used some methods to

detect the joint space, such as landmark detection, and segmentation analysis. Through careful

observation, we can see that the joint space and tissue bones have a significant boundary. That is

to say, the pixel values of these two groups are significantly different. Therefore, we decide to use

the K-means clustering method. The K-Means algorithm is an unsupervised clustering algorithm.

It is relatively simple to implement and has a good clustering effect, so it is widely used.

Compared with CNN, the K-means clustering algorithm has less calculation. Moreover, K-means

can achieve convergence faster.

Most important, the K-means algorithm has Stronger interpretability. However, the

problem for K-means is that the value of centroid is hard to decide. Nevertheless, it can be

solved: By giving a suitable value to k at the beginning, a clustering center is obtained by a K-

means algorithm. For the obtained cluster centers, the clusters with the closest distance are

merged according to the obtained distance of the k clusters, so the number of cluster centers

decreases. When it is used for the next cluster, the corresponding number of clusters also

decreases Smaller, and finally get a suitable number of clusters. A judging value E can be used to

determine the number of clusters to get a suitable position to stop without continuing to merge

cluster centers. Repeat the above loop until the evaluation function converges, and finally get the

clustering result of better clustering number. Besides, the K-means algorithm is susceptible to

noise value. In our knee data, noise does exist. According to our results, we know that most

values of the joint space are separated around 170. Meanwhile, there are hundreds of pixels

separated outside of the joint space. So, remove these noise values is still a big problem for our

model.

Until now, there are still many works can do to optimize this model. Our purpose for

using the clustering method is to detect the cluster accurately. It is possible to find a way to detect

the edge of the joint space. Then, making the pixel on both sides of the edge has a significant

difference. K-means is just one of the many clustering methods we used in this project. Many

advanced clustering methods, such as mean shift clustering (Nedrich, 2015), and Gaussian

mixture model (Carrasco,2020) can be used to do this process.

6.2 Shortcomings of Clustering Method

The idea we used in this paper is entirely new. So, there is no doubt that there are still

many shortages that exist in this paper. For instance, the values of pixel points around joint space

and bones are changed gradually. We cannot guarantee that the clusters are precisely the related

part we want in the X-ray image. So, the classification process does have errors. For some distinct

images, there is no significant difference between the joint space and bones the clustering method

could get a bad result; the mean and standard deviation of the cluster is not precise for these

images. The error is essential, especially for classify adjacent grades. That is the reason that the

existence of level 2 images could significantly influence the accuracy. So, to accurately detect the

clusters, it is necessary for us to do some processes to the edge and surrounded pixel values.

Image edges are one of the most basic features of an image, and often carry most of the

information of an image. So, detecting the edge and enhance the significance of the detected

edge. Then enhance the value of the point of one side and weaken the value of the other side. This

will be a very complicated process. Because we have thousands of samples. And each of the data

are got from different environment. The brightness, size of the bones can affect the edge

detection. We tried a couple of edge detection method such as Robert, Prewitt and Sobel operator

to detect the edge (Chapel et al., 2015). All of the three method are easy to be coded by python.

Among them, the Sobel operator has a relatively good result. The processed image is shown

below:

Figure 11
Edge Detection by Sobel Operator
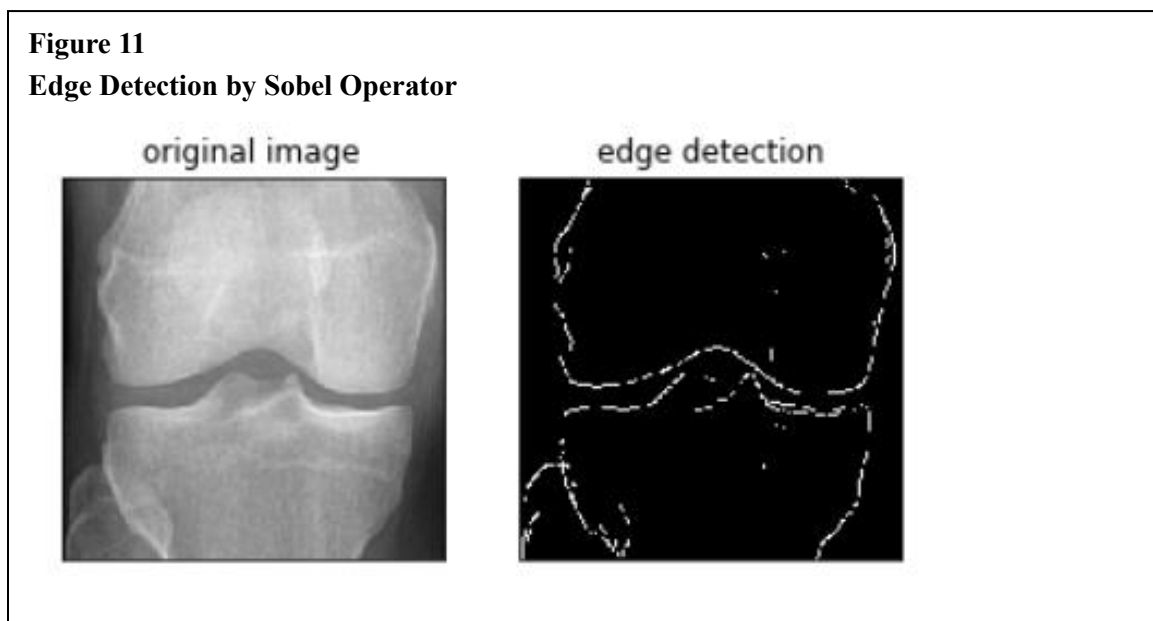
According to the images compared above, we can see that the Sobel operator did an

excellent job. However, this leads to a new problem. From the right image in figure 11, although

most of the edge pixels are detected, some noises still exist. Because what we want is to decay

and enhance the value of points around the edge. So, we need a complete, precise detection of the

edge. However, we see there is random noise in the image. Removing the noise data is one of the most necessary processes for using this method. It is possible that we can eliminate the noise data by adjusting the coefficients of the Sobel operator. Whereas the different condition of each x-ray images has different coefficients to be detected. To resolve this problem, we still have much work to do in the future.

What is more, we manually cut the image into a specific size. However, an automatic detection of the tissue and joint would be better. That is still related to edge detection. So, our method is limited to cutting images and divide it into a similar size for comparison, because of the difference in age of size of different patients. We cannot guarantee that all the images have the same size. That is to say, if we want to use the calculated area of joint space for detection, the discrepancy of knee size ratio has a severe effect on the results. Finding a way to change all cut images into the same size proportionally will improve accuracy. Noises also have many different types. We have to first detect what kind of noise it is, such as addictive noise, quantitative noise, multiplication noise, and so on (Luo, 2006). Among the existing denoising algorithms, some denoising algorithms achieve excellent results in low-dimensional signal image processing but are not suitable for high-dimensional signal image processing. Some of them do not have a good denoising effect, and lose some image edge information, or devoted to research and detection of image edge information, preserving image details. In our problem state in this paper, we not only need an accurate denoise algorithm; meanwhile, we still need to keep the edge. There is still much work that can do in the future.

Furthermore, we built a deep convolutional neural network and acquired a 67% accuracy. As we have mentioned before in this paper, the deep learning method has been prevalent for

detecting the knee OA. And many works of literature have worked on this. A pre-processing is

general for most of these papers. However, what we have done to the Xception is, the normalized

X-ray image is directly used as the input. A proper pre-processing for the input data is indeed

necessary for most CNN. But we did not add this process in our neural network. Comparing CNN

with the clustering method, we did not do multi-classification for our model. But we do binary

classification, which we can screen the test data from 0 to 4 one by one. According to table 4. We

can see that this method works. The problem is that it still needs more improvement.

7 Conclusions

In our paper, we used machine learning tools that have never been used in detecting knee OA. The method we used is: Using the clustering method (especially K-means clustering) to process the raw data and extracting 5 feature data from each image. So, the five features can constitute a five-dimensional vector for every single data. Therefore, the severity detection problem becomes a problem of classifying five-dimensional vectors. Our method works on the OAI dataset. Instead of multi-classification, we use binary classification to detect each specific severity gradually. The experiments show that our method has a relatively good result. The method we used gives new thinking in detecting knee OA. We did not use landmark detection or computer-aided analysis. Our method focusses on extracting the cluster rather than detecting landmarks. There are still many parts of our method that we need to improve. It is also great to combine our method with other existing methods to improve accuracy. We are interested in optimizing our method.

8 BIBLIOGRAPHY

Aljaaf, A. J., Hussain, A. J., Fergus, P., Przybyla, A., & Barton, G. J. (2016). Evaluation of machine learning methods to predict knee loading from the movement of body segments. *2016 International Joint Conference on Neural Networks (IJCNN)*. doi: 10.1109/ijcnn.2016.7727882

Antony, J., Mcguinness, K., Moran, K., & O'Connor, N. E. (2017). Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity Using Convolutional Neural Networks. *Machine Learning and Data Mining in Pattern Recognition Lecture Notes in Computer Science*, 376–390. doi: 10.1007/978-3-319-62416-7_27

Bleicher, A. (2017, August 9). Demystifying the Black Box That Is AI. Retrieved from https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/

Brahim, A., Jennane, R., Riad, R., Janvier, T., Khedher, L., Toumi, H., & Lespessailles, E. (2019). A decision support tool for early detection of knee OsteoArthritis using X-ray imaging and machine learning: Data from the OsteoArthritis Initiative. *Computerized Medical Imaging and Graphics*, *73*, 11–18. doi: 10.1016/j.compmedimag.2019.01.007

Carrasco, C. (2020, February 21). Gaussian Mixture Models Explained. Retrieved from https://towardsdatascience.com/gaussian-mixture-models-explained-6986aaf5a95

Chaple, G. N., Daruwala, R. D., & Gofane, M. S. (2015). Comparisions of Robert, Prewitt, Sobel operator based edge detection methods for real time uses on FPGA. *2015 International Conference on Technologies for Sustainable Development (ICTSD)*. doi: 10.1109/ictsd.2015.7095920

Chen, C., Xie, W., Franke, J., Grutzner, P., Nolte, L.-P., & Zheng, G. (2014). Automatic X-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements. *Medical Image Analysis*, *18*(3), 487–499. doi: 10.1016/j.media.2014.01.002

Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi: 10.1109/cvpr.2017.195

Darlow, B., Brown, M., Thompson, B., Hudson, B., Grainger, R., Mckinlay, E., & Abbott, J. H. (2018). Living with osteoarthritis is a balancing act: an exploration of patients' beliefs about knee pain. *BMC Rheumatology*, *2*(1). doi: 10.1186/s41927-018-0023-x

Du, Y., Almajalid, R., Shan, J., & Zhang, M. (2018). A Novel Method to Predict Knee Osteoarthritis Progression on MRI Using Machine Learning Methods. *IEEE Transactions on NanoBioscience*, *17*(3), 228–236. doi: 10.1109/tnb.2018.2840082

Gandhi, R. (2018, July 5). Support Vector Machine - Introduction to Machine Learning Algorithms. Retrieved from https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

Gorriz, M., Antony, J., McGuinness, K., Giro-i-Nieto, X., & O'Connor, N. E. (2019). Assessing Knee OA Severity with CNN attention-based end-to-end architectures. *Proceedings of Machine Learning Research*, 197–214.

Henche, H.-R. (1980). Approaches to the Knee Joint. *Arthroscopy of the Knee Joint*, 27–28. doi: 10.1007/978-3-662-02316-7_11

Hinton, G. E. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, *313*(5786), 504–507. doi: 10.1126/science.1127647

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *2015 Proceedings of the 32nd International Conference on Machine Learning*, 448–456.

Jackson, J. L., Omalley, P. G., & Kroenke, K. (2003). Evaluation of Acute Knee Pain in Primary Care. *Annals of Internal Medicine*, *139*(7), 575. doi: 10.7326/0003-4819-139-7-200310070-00010

Javaid, M. K., Kiran, A., Guermazi, A., Kwoh, C. K., Zaim, S., Carbone, L., … Nevitt, M. (2012). Individual magnetic resonance imaging and radiographic features of knee osteoarthritis in subjects with unilateral knee pain: The Health, Aging, and Body Composition Study. *Arthritis & Rheumatism*, *64*(10), 3246–3255. doi: 10.1002/art.34594

Kohn, M. D., Sassoon, A. A., & Fernando, N. D. (2016, August). Classifications in Brief: Kellgren-Lawrence Classification of Osteoarthritis. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4925407/

Lewis, J., & Carey, R. (2012). Surgery for Osteoarthritis of the Knee. *Osteoarthritis - Diagnosis, Treatment and Surgery*. doi: 10.5772/26904

Li, J., Liu, K., & Huang, Q. (2016). Utilizing Cloud Computing to Support Scalable Atmospheric Modeling. *Cloud Computing in Ocean and Atmospheric Sciences*, 347–364. doi: 10.1016/b978-0-12-803192-6.00017-7

Lim, J., Kim, J., & Cheon, S. (2019). A Deep Neural Network-Based Method for Early Detection of Osteoarthritis Using Statistical Data. *International Journal of Environmental Research and Public Health*, *16*(7), 1281. doi: 10.3390/ijerph16071281

Lindner, C., Thiagarajah, S., Wilkinson, J., Consortium, T., Wallis, G., & Cootes, T. (2013). Fully Automatic Segmentation of the Proximal Femur Using Random Forest Regression Voting. *IEEE Transactions on Medical Imaging*, *32*(8), 1462–1472. doi: 10.1109/tmi.2013.2258030

Luo, W. (2006). An efficient detail-preserving approach for removing impulse noise in images. *IEEE Signal Processing Letters*, *13*(7), 413–416. doi: 10.1109/lsp.2006.873144

Medina, S. (n.d.). Knee osteoarthritis : diagnoses, management and health effects. Retrieved from https://www.worldcat.org/title/knee-osteoarthritis-diagnoses-management-and-health-effects/oclc/958984751

Minciullo, L., Parkes, M. J., Felson, D. T., & Cootes, T. F. (2018). Comparing image analysis approaches versus expert readers: the relation of knee radiograph features to knee pain. *Annals of the Rheumatic Diseases*, *77*(11), 1606–1609. doi: 10.1136/annrheumdis-2018-213492

Moustakidis, S., Christodoulou, E., Papageorgiou, E., Kokkotis, C., Papandrianos, N., & Tsaopoulos, D. (2019). Application of machine intelligence for osteoarthritis classification: a classical implementation and a quantum perspective. *Quantum Machine Intelligence*, *1*(3-4), 73–86. doi: 10.1007/s42484-019-00008-3

Nedrich, M. (2015, May 27). Mean Shift Clustering Overview. Retrieved from https://spin.atomicobject.com/2015/05/26/mean-shift-clustering/

Oo, W. M., Hunter, D. J., & Linklater, J. (2017). Imaging in knee osteoarthritis.

Rad, A., Rahim, M. S. M., Rehman, A., Altameem, A., & Saba, T. (2013). Evaluation of Current Dental Radiographs Segmentation Approaches in Computer-aided Applications. *IETE Technical Review*, *30*(3), 210. doi: 10.4103/0256-4602.113498

Shamir, L., Ling, S., Scott, W., Hochberg, M., Ferrucci, L., & Goldberg, I. (2009). Early detection of radiographic knee osteoarthritis using computer-aided analysis. *Osteoarthritis and Cartilage*, *17*(10), 1307–1312. doi: 10.1016/j.joca.2009.04.010

Shamir, L., Orlov, N., Eckley, D. M., Macura, T., Johnston, J., & Goldberg, I. G. (2008). Wndchrm – an open source utility for biological image analysis. *Source Code for Biology and Medicine*, *3*(1). doi: 10.1186/1751-0473-3-13

Simonyan, Karen, Zisserman, & Andrew. (2015, April 10). Very Deep Convolutional Networks for Large-Scale Image Recognition. Retrieved from https://arxiv.org/abs/1409.1556

The Osteoarthritis Initiative a multi-center observational study of men and women. (n.d.). Retrieved from https://oai.epi-ucsf.org/datarelease/

Tiulpin, A., Melekhov, I., & Saarakkala, S. (2019). KNEEL: Knee Anatomical Landmark Localization Using Hourglass Networks. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. doi: 10.1109/iccvw.2019.00046

Tsai, M.-J., & Tao, Y.-H. (2019). Machine Learning Based Common Radiologist-Level Pneumonia Detection on Chest X-rays. *2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS)*. doi: 10.1109/icspcs47537.2019.9008684

Xue, N., Doellinger, M., Ho, C. P., Surowiec, R. K., & Schwarz, R. (2014). Automatic detection of anatomical landmarks on the knee joint using MRI data. *Journal of Magnetic Resonance Imaging*, *41*(1), 183–192. doi: 10.1002/jmri.24516

Zhang, Y., & Jordan, J. M. (2010, August). Epidemiology of osteoarthritis. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20699159/