Automated Detection For The Severity Of Knee Osteoarthritis From Plain Radiographs Using Machine Learning Methods

Aamir Yousuf Bhat, A.Suhasini,

Abstract— Knee joint pain is one of the foremost imperative clinical and debilitating features of osteoarthritis. It may be a late manifestation of osteoarthritis (OA) whose early indication includes joint space narrowing, bone subluxation, subchondral sclerosis, the formation of osteophytes and cartilage degradation. Even though the formation of osteophytes and joint space reduction is the early indications of the knee osteoarthritis (OA). The discovery is based on the Kellgren and Lawrence (KL) classification grades, which compares the distinctive stages of OA. Since the treatments for progressed radiographic knee OA are restricted, clinicians confront a significant challenge of distinguishing patients, who are at a high chance of getting affected by OA in a timely and effected way. Therefore we have created a straight forward self-assessment scoring framework for knee OA. OA is generally analyzed by specialists through manual assessment of patient's medical images, which are typically gathered in medical clinics. Checking the occurrence of OA to some degree is tedious for patients. In the expansion, the current studies are centred on consequently identifying OA through image-based machine learning algorithms. A right forecast of OA is a fundamental advance to successfully analyse and avoid the severity of OA. The objective of this study is to develop a machine vision approach for investigation of the severity of knee OA. The computation involves LBP (Local Binary Pattern) and PCA (Principal component Analysis) method. The processed PCA with LBP features are computed using FFNN (Feed Forward Neural Network), Multi-SVM and DBN with RBM (Deep Belief Neural Network with Restricted Boltzmann's Machine) classifiers for evaluating the knee osteoarthritis based on the Kellgren Lawrence grading system. The x-ray images used for the method are taken in stand Posterior Anterior (PA) view. The X rays images are used to obtain a region of interest (Extraction). The ROI extraction images undergo PCA and LBP feature extraction process. The texture features are then calculated from the ROI extraction. The extracted features are given to the ANN, Multi-SVM and DBN with RBM classifiers to classify the images to their grades. The highest accuracy rate of 83% is achieved by the DBN with RBM classifier for the classification of the images into the Grade 0, Grade 1, Grade 2, Grade 3, and Grade 4 osteoarthritis (OA). The goal of this paper is to analyse the severity of the osteoarthritis by using the plain x-ray radiographs.

Index Terms - Osteoarthritis, Local Binary Pattern (LBP), Principle component analysis (PCA), Knee x-rays Images, Multi-SVM, FFNN (Feed Forward Neural Network), DBN (Deep belief Neural Network), RBM (Restricted Boltzmann's Machine).

1 INTRODUCTION

T Osteoarthritis is one of the most forms of chronic disease mainly affects the knee joint, including joint space narrowing and degradation of cartilage. The knee is the most significant and most complex joint of the human body. It is a significant weight-bearing joint made up of condyles of tibia, femur and back surface of the patella. Cartilage provides the cover to the femur and tibia bone. It is a degenerative disease which occurs when the cartilage gets soft and gets damaged due to persistent wear and tear movements with ageing. According to the world health organization (WHO), about 80% of the population over the age of 65 years has been affected by OA [1]. Women's over the age of 55 are more likely to be affected by OA than men. Articular cartilage is the leading cause of the formation of osteoarthritis (OA). Roughening and diminishing of cartilage happens due to Osteoarthritis. This causes the stiffness in the cartilage and promptly gets wear with the increasing age, the ability of the cartilage recuperate from wounds goes on decreasing. The additional development of the bone at the outward side is called osteophyte due to which joint looks knobby. Due to the additional secretion of fluid causes the joint to look swollen, which causes severe pain and inability in the knee movement due to the stiffness. The expanding predominance of knee osteoarthritis implies that there is a developing requirement for useful clinical and logical devices to analyse the knee osteoarthritis in the early stage and to evaluate the severity in the progressive stages [2] [3]. Identifying the knee OA and evaluating the severity of knee OA are crucial for pathology, clinical decision making, and predicting the progression of the disease [4]. The reduction of the joint space and the formation of the osteophytes are the key obsessive features of osteoarthritis [5], which are easy to visualise utilising x-ray radiographs. There are few scores to measure the degree of osteoarthritis progressive stages using x-ray radiographs employing a standard score gets more chances to find a prepared radiologist. The Kellgren and Lawrence (KL) score had introduced their appraisal scale for knee OA [6], this has been adapted as a standard in many numerous health care systems and OA studies around the world. One of the primary factors utilize to measure the degree of PA is to calculate the joint space width (JSW) this estimation makes a difference to see the density of the cartilage and the formation of the osteophytes. The two most vital scores are provided to be effective within the conclusion and the classification of the disease. To analyse osteoarthritis, medical imaging may be exceptionally a vital tool utilizing different radiological methods such as x-rays or MRI interpretation is conceivable to get an excellent approach of the OA stages. Although in some

works [7][8] it is possible to get superior outcomes using Magnetic resource imaging (MRI) as compared to x-ray imaging, but this is often since the nature of the image itself.

The clinical reviews have confirmed that knee harm could be a prime pointer for the advancement of knee osteoarthritis. The connections between the joint damage and knee osteoarthritis (OA) were at first delineated by Kellgren and Lawrence. The Kellgren and Lawrence grading system are most affirmed system for the systemization of single joint into five grades [9]. Any unclear conclusion drawn from x-ray images for osteoarthritis makes treatment troublesome and faulty. In this manner, to make symptomatic procedure increasingly proficient and dependable, it is required to build up a machine vision diagnostics for early identification of osteoarthritis and prediction of the disease in the beginning period. The more precise identification will prompt progressively compelling treatment for anticipating further damage to fragile tissue and tendon. The aim of the present work is to consolidate the machine vision approach that prompts more precision in the examination of the illness to some degree.

The last stage of OA can cause the cartilage to wear out near almost totally. Due to this the adjoining bones of the joint rub with each other. Other components that can cause osteoarthritis are heredity, obesity [10], overs use of the joint and tall bone mineral thickness. There are diverse ways in which an individual gets influenced by OA. It causes musculoskeletal torment and disability within the knee joint [11]. The objective of this paper is to create a strategy for the investigation of OA for the knee joint. The paper is organized into the five sections. Section I gives the introduction. Section II discusses the related work. In the third one, it describes the methodology used for acquiring the knee images. In section IV the actual proposed method used different steps to be carried out like Feature Extraction and then classification for the grading according to the features calculated. Section VI discusses experiments and results which are at that point followed by conclusion and references.

2. Related Work

The related work reveals that various techniques have been utilized for the acknowledgement and classification of osteoarthritis. In the different literature, computer-assisted and semi-computerized techniques have been proposed to examine the diseases utilizing x-ray imaging for knee osteoarthritis.

Hegadi et al. [12] used a block-based texture analysis approach to identify the synovial cavity region and support vector machine (SVM) classifier is adapted to classify the images. Nine equally sized blocks of single inputted images are considered to extract four different texture features (skewness, kurtosis, standard deviation and energy). In total 36 features are extracted and stored in feature vector to achieve the desired accuracy. The proposed method demonstrated 80% and 86.7% accuracy for the healthy and affected images respectively

Anifah et al. [13] proposed a classification technique based on the self-organised map. In the proposed work, the limited adaptive histogram (CLAHE) equalisation is used to increase the contrast of the image. The decision of whether an image is of the left or right knee is based on the template matching technique. To identify the region of the synovial cavity, they segmented the image of the knee using Gabor's kernel, template matching, the row sum chart and the grey level centre of the mass method. The characteristics of the gray level matching matrix (GLCM) are used for additional classification of the data into five degrees of OA. They concluded that the accuracy of the classification of images with grade 0 and 4 osteoarthritis is high than images that have grade 1 to grade 3. They used three characteristics such as accuracy; Specificity and sensitivity for research evaluation.

Brahim et al. [14] applied machine learning to identify knee osteoarthritis in its initial stage using knee x-rays. Most of the previous studies focused on image data with machine learning or deep learning with image data, such as magnetic resonance imaging or X-rays. However, as far as we know, prediction studies that use statistical data with machine learning or deep learning are rare.

Lior et al. [15] et al. in 2008 he developed an automated approach to the detection of OA according to the Kellegren-Lawrence method that shows the probability of OA at different stages, that is, KL grade 0,1,2,3. His work on 20 preselected images, each image is 150 * 150 joint central windows, and then these images were reduced by a factor of 10 in 15 * 15 images.

Tati et al. [16] propose a technique to automatically determine the region of interest for the evaluation of knee osteoarthritis to horizontal and vertical translation. The detection of boundaries between the femur and the joint space, for some image boundaries, appears discontinuously.

Aleksei Tiuplin et al. [17] have proposed an algorithm for localisation of the knee joint area for the assessment of X-ray radiographic images of the knee. The overall work is divided into two parts: it automatically locates the joint area and calculates the calculation using the HOG method. The outcome shows that this technique is appropriate for large-scale investigations. The results are also unilateral, and the method is of higher computational complexity. It has been proposed to improve execution by using reduced-scale images for joint detection and then the detected scaling region for further analysis

Pooja P et al. [18] actively used the distance-based model to calculate the radiographic parameter associated with osteoarthritis in knee radiographs. The characteristics are calculated using the local binary configuration technique and are characterized in KL degrees using the K-NN classifier. Experimental results showed a classification accuracy of 93.8%

for KL-0, 70% for KL-1, 4% for KL-2, 10% for KL-3 and 88.9% for KL-4. As a result, the LBP approach produces better results and provides better quality for the early detection of OA.

Sreepama Banerjee et al. [19] have portrayed a class of algorithms *that* depend on cellular networks for identifying OA based on hand x-ray images. Cellular neural network algorithms are valuable when quick and powerful pre-processing is required. The parameter utilized for the identification of OA is osteophytes or bone spurs. The algorithm uses a standard convolution neural network (CNN) templates and spicules seeker algorithm to isolate the osteophytes. They have shown the result with a 90% accuracy rate

G. W. Stachowiak et al. [20] considered trabecular bone textures as a major region for experimenting. The assessment of osteoarthritis is done by creating an automatic decision frame using radiological images of the hands and knees. The created framework is suitable for the investigation of radiological images of the hands and knees. The authors hypothesised that the technique used differentiated not only osteoarthritis and healthy knee joints, but also pre-radiographic bone changes in osteoarthritis. In the future, the authors might consider the regression model for OA prediction and try to use this model to examine other conditions such as rheumatoid arthritis and osteoporosis.

3. Materials and Methods

The proposed method for the classification of the knee osteoarthritis in x-ray images is developed based on LBP, PCA feature extraction and by applying to FFNN, multi-SVM and DBN with RBM classification for building the classifier.

3.1 Data Set

We procured the information from the data sets taken from various hospital and other clinics. We have utilized the entire cohort of the dataset for preparing our model and the pattern from the dataset for testing. The entire dataset contains the information from 126 subjects, which includes data for both men and women aged in between 50-79 years old. The images in the data set are graded according to semi-quantitative KL grade system, which is categorized into five stages. Grade-0 or KL-0 (Normal, No OA), KL-1 (Doubtful), KL-2(Early OA changes), KL-3 (Moderate OA), KL-4 (End stage, severe OA). We prepared all our models for the (right and left knee) and utilize to the horizontally flipped clear out ones to extend the

dataset size. The description for the OA grades is given in Table 1.

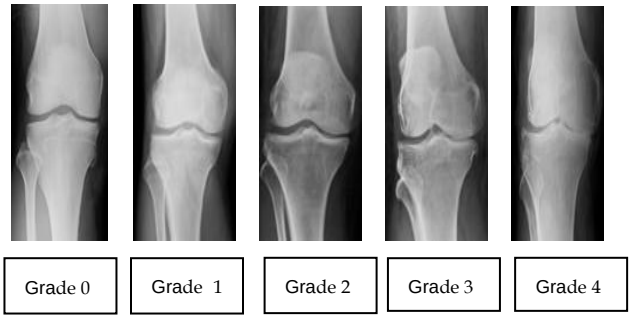


Fig 1 Stages of Osteoarthritis

Figure 1 Posterior Anterior view (PA) of the knee joint with the Kellgren Lawrence grades of Grade 0, Grade 1, Grade 2, Grade 3 and Grade 4 of patients having Osteoarthritis (OA) with the presence of small subchondral sclerosis, formation of osteophytes, Joint space narrowing, deformity of the joint surface are typical for the diagnosis of Osteoarthritis (OA).

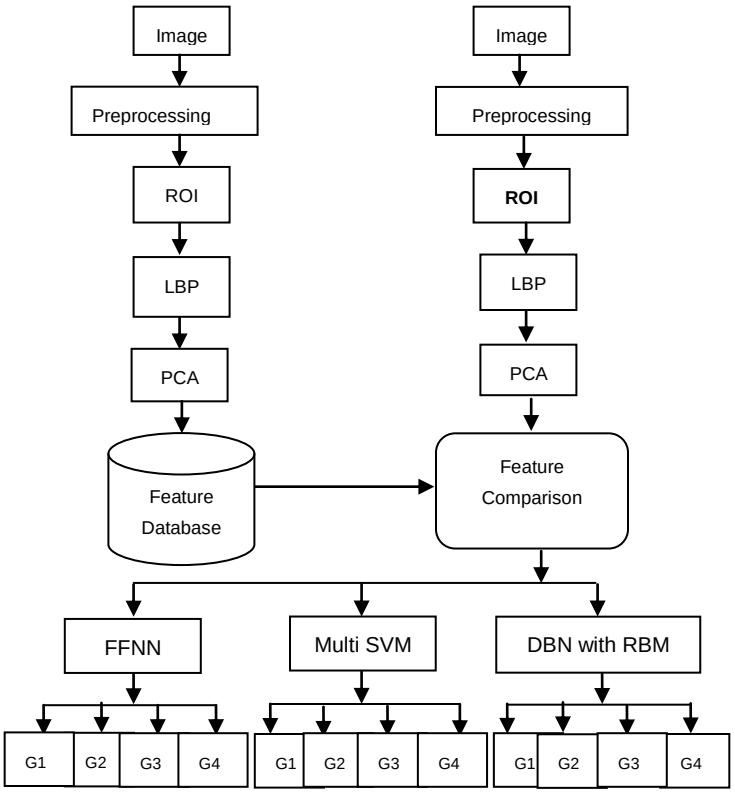


Fig. 2 Block diagram for the proposed methodology for OA (Grade 0, Grade 1, Grade 2, Grade 3, and Grade 4)

Table 1 KL Grading System

KL Grades	KL Description
Grade 0	No radiographic features of OA present
Grade 1	Doubtful OA, Narrow joint spacing and possible osteophytes lipping.
Grade 2	Mild OA, Definite osteophytes and definite narrowing of joint space width.
Grade 3	Moderate OA, Multiple osteophytes, definite narrowing of joint space, severe sclerosis, and possible deformity of bone contour.
Grade 4	Severe OA, Large osteophytes, marked joint space width, severe sclerosis, bone subluxation, bone angulation and definite deformity of bone ends.

3.2 Pre-processing

The main objective of pre-processing is to enhance the visualization of the image. Pre-processing mainly focuses on to evacuate the clamor, stabilising the intensity of the image and clear the artifacts. It is the technique of improving the image information earlier to computational processing. Pre-processing is carried out in one of the following forms:

Image re-sampling- It is a procedure to convert an inspected image from one coordinate to another. Utilizing the mapping work of the dimensional transformation, the two facilitate frameworks are related to each other. The reverse mapping work is connected to the yield pixel so that the obtained re-sampling pixel is switched to get the initial input pixel.

Contrast Enhancement- In order to make the image more reasonable for positive applications, different upgrade (contrast enhancement) must be utilized. It improves the permeability and the transparency of the image. When the image values of the low contrast images are extreme, the contrast enhancement extends the intensity of the pixel. The images can have reduced measurable range either due to the low quality of the imaging devices or the excellent outside condition during the acquisition process. By the different differentiate enhancement techniques; histogram adjustment method is broadly utilized since of its use and effectiveness. The method of histogram equalisation is to extend the intensity of the input image to make indistinguishable dissemination within the goal that the dynamic scope of the image is completely discouraged.

Fig. 3 (a) Original Image 3 (b) Pre-processed Image

3.3 Region of Interest (ROI) Extraction

ROI portion is extracted by localising region-based active contours algorithm. It is a model-based technique, and it is insensitive to noise. It is not based on global region models; as an alternative, it allows the foreground and background to be depicted in terms of smaller local regions, removing the assumption that the foreground and background regions can be represented with global statistics. Local regions are leading to the structure of a set of local energies at each point along the curve. To optimize the local energies, each point is measured separately and moved to minimize (or maximize) the energy calculated in its local region. To estimate these local energies, local neighbourhoods are split into local interiors and local exteriors by the evolving curve. The energy is then optimized by fitting a model to each local region. Our proposed model implements segmentation by minimizing the following energy function

$$E(\emptyset) = \int \Omega_x \delta(\emptyset(x)) \|\nabla \emptyset(x)\| dx + \int \Omega_x \delta(\emptyset(x)) \int \Omega_y K_\sigma(x,y) F_{new}(I(y), \emptyset(y)) dy dx \quad (1)$$

Where, image domain is represented by Ω, x and y are independent spatial variables, and the smooth Dirac delta function, δ, is defined as,

$$\delta_{\emptyset(x))} = \begin{cases} 1, & \emptyset(x) = 0 \\ 0, & |\emptyset(x)| < \epsilon \\ \frac{1}{2\epsilon}\left\{1 + cos\left(\frac{\pi \emptyset(x)}{\epsilon}\right)\right\}, & Otherwise \end{cases} \quad (2)$$

IV Feature Extraction

4.1 Local Binary Pattern (LBP)

Local Binary Pattern is an image operator which changes an image into an image of integer labels depicting small scale appearance of the image. These labels are mainly utilized to assist the further investigation. LBP employs the grey level structure of the image. The design works on a 3×3 window over the selected portion of the image. The centre pixel value is compared with the neighbouring pixel values over the window. A single pixel is compared with the eight distinctive pixels over the window 2^8=256 different designs can be obtained for the selected portion of the image. The LBP

operator is given by:

$$LBP(X_c, y_c) = \sum_{n=0}^{7} 2^n g(I_n - I(X_c, y_c) \quad (3)$$

Where I_c is corresponding to the centre pixel in the window and I_n are the eight pixels surrounding the centre pixel. LBP (x_c, y_c) is the value at the centre pixel (x_c, y_c), I_n and $I(x_c, y_c)$ are the values of the neighboring pixel and centre pixel, respectively.

$$Sx = \begin{cases} 0, x < 0 \\ 1, x > 0 \end{cases} \quad (4)$$

The LBP operator works as follows:

$$\begin{bmatrix} 77 & 59 & 42 \\ 50 & 50 & 45 \\ 67 & 21 & 3 \end{bmatrix} \xrightarrow{\text{Threshold}} \begin{bmatrix} 1 & 1 & 0 \\ 1 & & 0 \\ 1 & 0 & 0 \end{bmatrix}$$ Binary code 1100011

Fig. 4 Operation of LBP (Local Binary Pattern)

It is a typical combination of grey scale-invariant with a presence of certain condition measure that is by computing the difference of the average grey level of that neighbourhood.

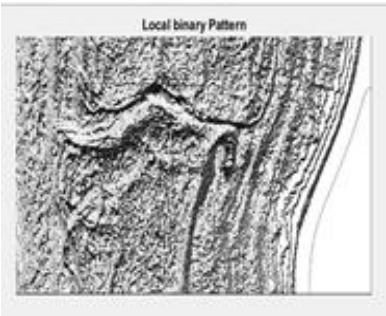


Fig. 4 Visualization of Local Binary Pattern

It serves as a descriptor of nearby spatial pattern and grayscale contrast. LBP label may be a paired number made for each pixel, where each bit is allocated based on its differences from one of the pixels at a given radius. Figure 1 demonstrates a basic form that name every pixel by a number got from its 3 neighbourhoods. Neighbouring pixels are allocated a binary value of 1 if more prominent than the inside pixel and 0 generally. LBP is an operator for a surface depiction that is based on the signs of contrast between neighbour pixels and centre pixels [21] [22].

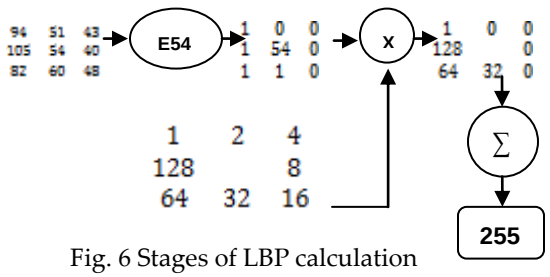


Fig. 6 Stages of LBP calculation

Fig.4 shows the calculation of the LBP values. For each pixel in the image, a binary code is obtained by thresholding its neighbourhood with the value of the middle pixel. The neighbouring pixel will get the value of 1, if the pixel value is less than the threshold. The histogram will be developed to determine the recurrence values of binary patterns. The number of histogram bins depends on the number of included pixels in LBP calculation. If LBP employs 8 pixels, the number of histogram bin will be 8 or equal to 256. The basic version of the LBP operator employees the centre pixel value as the limit to the 3×3 neighbouring pixels. The thresholding operation will make a binary pattern representing surface characteristic.

Other expansion to the initial LBP operator is uniform pattern LBP, which can be utilized to decrease the length of the feature vector and execute a straightforward rotation-invariant descriptor. This expansion is propelled by the way that some binary pattern happens more commonly in facial imaging than others. A neighbouring binary pattern is called uniform on the off chance that the binary pattern contains at most two bitwise moves from 0 to 1 or vice-versa when the bit design is navigated circularly. The output value is diminished when LBPP, R is mapping to LBP^{U}_{P, R^2} based on the equation$P(P-1)+3$. The normal, which is 256 bins pattern are when utilizing (8, R) neighbourhood will get to be 59 patterns [23].

The Uniform pattern is mapped by the rotation invariance of LBP to include vital texture information to the image given as:

$$LBP^{riuz}_{P, R} = \begin{cases} \sum_{P=0}^{P-1} s(g_p - g_c), if\ (LBP_{P R}) \leq 2 \\ P+1 \end{cases} \quad (5)$$

4.2 Principal Component Analysis

Principal component analysis (PCA) is a dimensionality reduction method that's regularly utilized to diminish the dimensionality of the large data sets by changing a large number of factors into a smaller one that still contains most of the data in the data set. By diminishing the number of a factor of the information set comes at the cost of precision, but the trick in dimension reduction is to exchange a small accuracy for simplicity [24] [25]. Since little information sets are less demanding to investigate, visualize and to make analysing information much more straightforward and quicker for machine learning algorithms without essential variables to process.

PCA in signal processing can be depicted as a change of a given set of n input vectors with the same length k shaped with the n-dimensional vector $x = [x1, x2, x3 \ldots\ldots xn]^T$ into a vector

$$y = A(x - m_x) \quad (6)$$

This perspective empowers to perceive a straightforward equation, but it is essential to keep in mind that each row of the vector x compromises of k values belonging to one input. The vector m_x in eq. 1 is the vector of mean estimation of all input variables defined by a relation.

$$m_x = E\{x\} = \frac{1}{k} \sum_{k=1}^{k} x_k \quad (7)$$

Matrix A in eq. 1 is determined by the covariance matrix c_x. Rows

in the A matrix are formed from the eigenvectors e of c_x ordered according to corresponding eigenvalues in descending order. The evaluation of the c_x matrix is possible according to a relation.

$$c_x = E\{(x - m_x)^T\} = \frac{1}{k} \sum_{k=1}^{k} x_k x_k^T - m_x m_x^T \quad (8)$$

As the vector x of the input variable is n-dimensional, it is evident that the size of c_x is n×n. The elements c_x (i, i) lying on its main diagonal are the variances of x and the other values.

$$C_x(i, i) = E\{(x_i - m_i)^2\} \quad (9)$$

C_x (i, j) determine the covariance between input variables x_i and x_j

$$C_x(i, j) = E\{(x_i - m_j)\} \quad (10)$$

The rows of A in eq. 1are orthonormal, so the inversion of PCA is possible according to the relation

$$x = A^T y + m_x \quad (11)$$

The kernel of PCA defined by eq. 14 has some interesting properties from the matrix theory, which can be used in the signal and image processing to full fill various goals.

V Classification

5.1 FFNN (Feed Forward Neural Network)

FFNN is the simplest type of ANN network in which the connection between units does not form a loop. They are called feed-forward neural networks since the data travels to the network from input nodes through hidden nodes and then pass to the output node. There is no input association in which the outcome of the framework is nourished back to itself. When FFNN is amplified to incorporate input associations, they are called recurrent neural networks. They are primarily used for supervised learning in cases where the information to be learned is neither successive nor time-dependent. That is for a classifier $y = f^*$ (x) maps an input x to a category y. FFNN defines a mapping $y = F(x; \theta)$ and learns the value of the parameter θ that results in the best approximate function. FFNN consists of three layers viz input layer, hidden layer and output layer. Backpropagation is the conventional preparing strategy for FFNN amid which the neuron adjusts their weights to procure unused information. The learning stage in FFNN with backpropagation happen during the preparing stage in which each input design from the preparing set is connected to the input layer and after that propagates it forward [26].

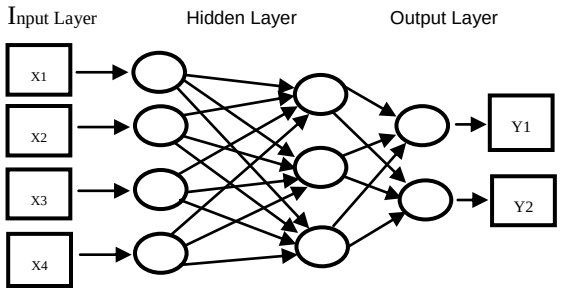


Fig 7 Feed Forward Neural Networks

- **Input Layer:** In the input layer, the different input parameters are passed for preparing. In the proposed method linear activation function is utilized in the input layer
- **Hidden Layer:** Hidden layer acts as an intermediate layer between the input layer and the output layer performs the middle computation of the framework. There may be single or numerous hidden layers with the number of hidden nodes changing from l to 2l, where li is the number of inputs within the input layer. In the worst-case, it is equal to the number of samples in the data set.
- **Output Layer:** In the output layer, it takes the data from the hidden layer and produces the final result of the framework.

In FFNN each of these layers may take one or number of artificial neurons. An interface exists between each neuron in one layer to another neuron in the next layer. That is the neuron in the input layer joins every other neuron with the hidden layer. Each neuron in the hidden layer will encourage connect to neurons in the next layer that follows this hidden layer which may be another hidden layer or an output layer depends on the nature of the architecture design. Furthermore, the activation function may exist between various layers that is in between the input layer and the hidden layer, between the hidden layer and the output layer these activation capacities may or may not be the same.

5.2 Multi SVM

SVM is a supervised machine learning algorithm that analyses data for regression and classification. SVM uses a strategy called kernel trick to convert the data and based on these changes, it finds an ideal boundary between the conceivable outputs. SVM works by mapping the information to a high dimensional feature space so that the information pints can be categorised indeed when the data is not something differently detachable. It employs nonlinear mapping to change the original information into a higher dimension. The objective of SVM is to construct a work, which can accurately anticipate the class to which the unused point belongs and to which the old points belong to. With a suitable nonlinear mapping, two

datasets can continuously be separated by a hyper plane. The primary purpose of SVM is to isolate the data with decision boundary and expand it to nonlinear boundaries utilizing kernel trick [27]. The advantage of SVM is flexible implies distinctive kernel function can be specified for the decision function.

For binary classification issues, the idea behind SVM is to separate the data in the most excellent strategy. It is utilized when we need to classify the two data sets. SVM has been primarily designed for two-class classification problems [28].

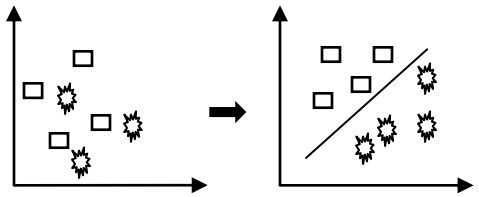


Figure 8(a) Nonlinear Separable 8(b) Linearly Separable

- Linearly Separable
- Non-Linear Separable

For linear separable, there are numerous linear decision boundaries that separate the information. Only one among them would accomplish the most extreme division. Suppose if we use a decision boundary for classification that may be closer to one set of datasets compared to others, in this way the concept of a hyper plane is used as an alternative solution. Support vectors can be depicted as those information points that the edge pushes up against [29]. The typical issue here is to find the only optimal margin of the separating hyperplane $w_x{}^t + b = 0$, the one that gives the greatest margin between classes.

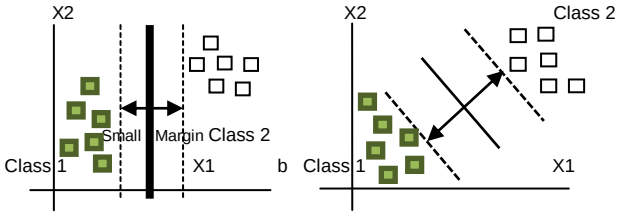


Fig. 9 (a) Small and Large Margin separation 9(b)

For the nonlinear separable the information is not linearly separable. Linear SVM can be amplified to produce nonlinear SVM's for classification of linearly indistinguishable data. Such type of SVM is equipped for finding nonlinear decision boundaries.

For multiclass classification, there are two approaches, namely direct and indirect approaches. For a direct approach, it includes Bayes classification, k nearest neighbour and decision tree. Multiclass classification through incorporate one-vs-one, one-vs-all, Direct acyclic, chart SVM and Error coding output codes. Closest neighbour classifier depends on closeness. For an obscure tuple, k nearest neighbour classifier looks the design space for the k preparing tuple that is closest to the new tuple. K preparing tuples are the k nearest neighbour of an unknown tuple. Closeness is characterized in terms of a distinct metric such as Euclidean distance. A closest neighbouring classifier can be significantly moderate when ordering the test tuples.

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (12)$$

The decision tree is a flow chart like structure where each inner node indicates a test on an attribute, each branch speaks to as a result of the test, and each leaf node holds a class name. In classification attribute values of the tuple are tried against the decision tree. The highest node in a tree represents the root node of that tree. The decision tree can be effortlessly changed into classification rules. These trees are prevalent, since it does not require any space information parameter setting and can handle multidimensional with quick speed and good accuracy [30].

Bayesian classification predicts the class membership probabilities such as the likelihood that a given tuple belongs to a particular class. It depends upon the Bayes Theorem. Bayes Theorem gives a way of calculating posterior probability given by:

$$P\left(\frac{H}{X}\right) = P\left(\frac{X}{H}\right)P(H)P(X) \quad (13)$$

Bayesian classifier has the least error rate in comparison to all other classifiers, but in practice that is not continuously, such as the need for accessible likelihood information. For multiclass classification in SVM, the thought of utilizing a hyperplane to isolate the information into two bunch sounds well when there were two target categories. Various approaches have been proposed, but there are two most common approaches portrayed below [31].

- One vs all
- One vs one
- Directed acyclic graph

In one vs all where each category is part out, and all of the other categories are combined and to select the class, which classifies the test information with the most noteworthy margin. It separates an m class problem into m binary issue. The learning step of the classifier is done by the complete preparing information considering the designs from the specific class as positive and all other illustration as negatives. In the approval stage, a designer is displayed to each one of the binary classifiers and after that classifier, which gives a positive yield shows the output class. The positive result is not unique in various classes, and some tie-breaking strategies are compulsory. The first familiar approach employs the confidence of the classifier to choose the outcome, predicting the class from the classifier with maximum certainty. Instead of having a scoring matrix when managing the resulting of

OVA (one vs all) classifier (i r in [0, 1] is the certainty for the class) a score vector is utilized.

$$R = (r1, r2, r3, \dots \dots rn) \quad (14)$$

Another procedure is to build a set for one vs one classifier by separating an m class problem into m (m-1)/2 binary problems. Each issue comes up with a twofold classifier, which is mindful of distinguishing between a distinctive pair of classes. The learning stage of the classifier is done utilizing as training information where a subset of occasions from the original data set which contains any of the two corresponding class names and the occurrences with different class labels are disregarded [32].

In the approval stage, the design is displayed to each one of the twofold classifiers. The yield of a classifier is given by r_{ij} in [0, 1] is the certainty of the twofold classifier segregating classes I and j in favour of the previous classes. Class with the most significant certainty is the outcome class of a classifier. The outputs are represented by a score matrix R given by:

$$R = \begin{matrix} - & r12 & r1m \\ r21 & - & r2m \\ rm1 & rm2 & - \end{matrix}$$

Direct acyclic graph can be utilized to represent a set of programs where the input, output or execution of one or more programs is subordinating on one or more other programs. The programs are nodes in the graph and the edges distinguish the dependencies. A coordinate acyclic graph may be a linear graph that contains no cycles.

5.3 Deep Belief Networks (DBN)

Deep belief network is defined as a state of continuous Boltzman machines in which each RBM layer communicates with both the previous and the subsequent layers. The node of any single layer does not communicate with each other along the side. This stack of RBM might end with a softmax layer to form a classifier. With the exemption of the first and final layer in a deep belief, network incorporates a twofold part; it serves as the hidden layer to the nodes that come before it and the visible layer to the nodes that come after it. Deep belief networks are utilized to recognize cluster and produce images, video arrangements and video capturing information. A continuous deep belief network is an expansion of a deep belief network that acknowledged a continuum of decimals instead of binary data. Restricted Boltzmann Machine unsupervised learning has the advantage of fitting the features of the samples, so when an output of the hidden layer in an RBM, we will utilize it as the visible layer's input of another RBM. This procedure can be regarded as further feature extraction from the extracted features of our samples. In figure 7 it is shown by utilizing the output of the upper RBM's hidden layer as the input of the lower RBM's visible layer.

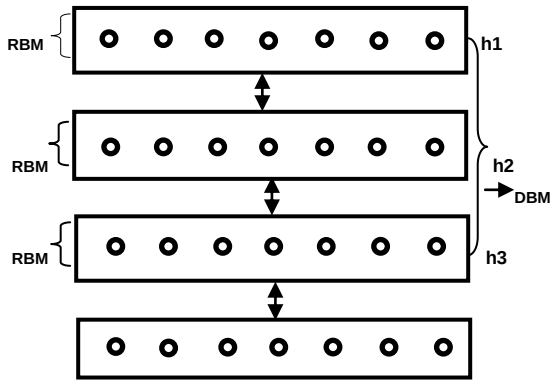


Fig 10 DBN Stacked by three RBM Machines

However, RBM's learning process is unsupervised learning, so the deep belief network can work without supervising. In the event, we need to utilize it as a classification. We should include another system of supervised learning, which can group the samples dependent on the features extracted by the DBN. Its primary thought is that DBN can extract the features of the samples as well. This will make the classifier work superior to that without DBN.

The goal behind training a deep belief network by training a sequence of RBM is that the model parameter θ learned by an RBM define both P(v/h, θ) and the earlier distribution over hidden layer output vectors P(h/θ), so the probability of creating a visible layer output as vector v can be defined as [33]:

$$P(v) = \sum h \; P\left(\frac{h}{\theta}\right).P\left(\frac{v}{h}, \theta\right) \quad (15)$$

After learning the parameters θ and P (v/h, q) is preserved, while P (h/q) can be replaced by an improved model, that is learning by treating the hidden activity vectors H= h as the training data for another RBM. This replacement progresses a variation lower bound on the probability for training data under the composite model. As per the experimental investigation of DBM, it could be utilized as a feature extraction method for dimension reduction in case of not utilizing class names and backpropagation in the DBM architecture [34]. Again, on the other hand, when connecting class names with feature vectors, DBN is utilized for classification. There are two types for DBN classification architecture; one is the Backpropagation DBN (BP-DBN), and the second one is the associated Memory DBN (AM-DBN). For both the designs, when the number of conceivable classes is huge, and the distribution of frequencies for different classes is distant from uniform, it may some of the time be advantageous to utilize a distinctive encoding for the class targets than the standard one of k-softmax encoding.

5.4 Restricted Boltzmann Machine

Restricted Boltzmann Machine (RBM) is a virtual based generative model that compromises of a layer of paired visible units v and a layer of binary hidden units h associated by symmetrical weight connections [35] [36] are shown in figure 5.

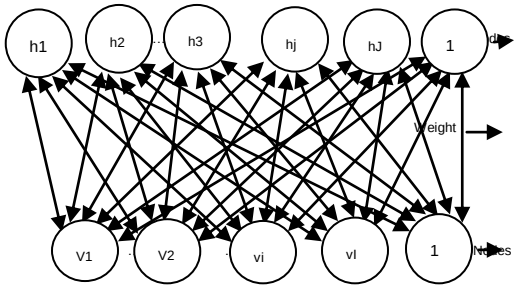


Figure 11 Schematic representation of RBM

It is an algorithm which is valuable for dimensional diminishment, regression classification, feature learning and point modeling. For an observed state, the energy of the joint step of the visible and hidden unit is given by

$$E(v,h) = \sum_{i=1}^{I} ai\ vi - bj\ hj - \sum_{i=1}^{J}\sum_{i=1}^{I} wji\ vi\ hj \quad (16)$$

In an observed state, where a_i represent the bias of the visible units I, b_j represent the bias for the hidden unit j, and w_{ji} represent the weight related to the connection between the visible unit I and the hidden unit j. The RBM assigns likelihood for each step (v_i h) utilizing the energy given by the Eq. 4.

$$P(h,v) = e^{-E(h,v)} \quad (17)$$

Where z is the partition function, determined by summing the energy of all possible (v, h) configurations

$$z = \sum_{v,h} e^{-E(v,\,h)} \quad (18)$$

$z = \sum v,h \exp(-E(v,h)$ Is the Boltzmann partition function. After summing all the conceivable hidden units, the marginal distribution over the visible units is created as:

$$P(v) = 1/Z \sum_h \exp(-E(v,h) \quad (19)$$

For a random input configuration v, the state of the hidden unit j is set to 1 with probability given by

$$P\left(hj = {}^{1}/_{v}\right) = \sigma(bj + \sum_{i=1}^{I} vi\ wji \quad (20)$$

The probability assigned to a visible vector v is given by

$$P(v) = \sum_k P(v,h) = \sum_k P\left(\frac{v}{k}\right) P(h) \quad (21)$$

Consequently, the likelihood given to a particular specific training vector v can be raised by the weights and the inclination of the network in arrange to bring down the energy of that specific vector while raising the energy of all the others. In this conclusion, we can perform a stochastic gradient ascent on the log-likelihood surface of the preparing information by processing the subordinate of the log-likelihood regards to the load of the network which is given by:

$$\frac{\partial \log(v)}{\partial wij} = (vi\ hj)\sigma - (vi\ hj)\infty \quad (22)$$

Where hi the desired about the subscript specified distribution and the second term is often approximated by contrast convergence calculation. The weight update standard would then be able to be depicted as:

$$\Delta wij = \epsilon\ (\langle vi\ hj \rangle\ data\ \ \langle vi\ hj \rangle recon) \quad (23)$$

Where ϵ is the learning rate and v_i^{neg} is a sample. For a given set of training data $\{v^1, v^2, v^3 \ldots vn^j$ the weight update function can be written as:

$$\Delta wij = \epsilon\ (P\left(\frac{hj}{vi(n)}\right) vi\ (n) - P\left(\frac{hj}{vi(n)neg}\right) vi(n)neg) \quad (24)$$

The update rules for both visible and hidden layer are:

$$bi = bi + \epsilon\ (< bi >) data - < bi\ neg >) \quad (25)$$
$$ci = ci + \epsilon\ (< vi > data - < vi\ neg >) \quad (26)$$

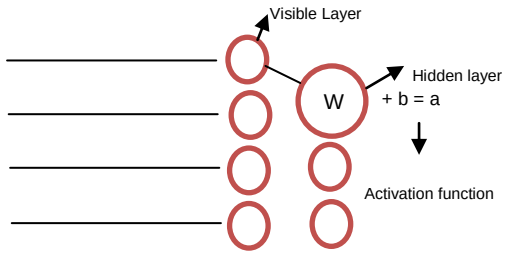
The behaviour of Restricted Boltzmann Machine can be clarified by altering the weights and affects to lower the energy on preparing the data [37].

Layers in RBM

The RBM is a shallow two layers network that establishes the structure of deep belief networks (DBM). The first layer of RBM is called the input layer or a visible layer. The second layer is called hidden or covered up layer. Each circle represents a neuron-like unit called a node. The hubs are associated with each other over layers, but no two hubs of the same layer are interconnected. The confinement in RBM is that there are no intralayer connections. Each node may be a locus for calculating that forms input and starts by making stochastic choices whether to transmit that input or not.

Working of RBM

Every visible node takes low-level components from a thing in the dataset to be learned. In a hidden layer for node 1 x is increased by weight and added to a bias. The consequent of those two tasks is encouraged into an initiation work, which delivers the nodes yield or the quality of the signal going through it given input x.



X- Input, W-Weight. A- Activation Function
Fig. 12 Working of RBM

Each x is increased by a particular weight, the items are summed included to an inclination, and once more, the result is passed through an actuation work to deliver the node's output.

VI. Experiment and Results

As we have mentioned, the experimental dataset contains the information from 126 subjects, and the experimental process is implemented in Matlab R2016 (a) with AMD A4-330 MX APU Radeon (tn) HD processor with 4 GB RAM. The experiment has been implemented over different knee X-ray images in the format of JPEG, from distinctive age group and sex. Images used in the work are taken from various hospital and other clinical centres. The images used for this experiment are taken in stand posterior anterior view. Local Binary Pattern (LBP) and Principle Component Analysis (PCA) have been used for feature extraction. Then feature extraction is classified using Multi SVM (Support Vector Machine), FFNN (Feed Forward Neural Network) classifier and DBN (Deep Belief Network) Classifier. The proposed method is given as:
Input x-ray image
Step 1: Pre-processing of the x-ray image, which includes noise removal, equalizing histogram and auto-brightness and contrast, has been adjusted over the given grey-scale images and it is resized to the dimension of 512 x 512 pixels.
Step 2: ROI (region of interest) segmentation is done by localizing region-based active contours algorithm.
Step 3: Feature extraction using Local Binary Pattern (LBP). Feature Reduction is done by PCA.
Step 5: Classification of feature values using Multi-SVM, FFNN and DBN classifiers.
Step 6: END.

The classification results are shown in the confusion matrices. From the table 2, it has been found that out of 126 knee x-ray images, 51 images have been used for training and 75 images have been used for testing purpose. The original input size of the x-ray image was 2916× 2460 dimensions with the horizontal resolution of 96 dpi and the vertical resolution of 96 dpi. The input image size of an image has been resized into 512 x 512 dimensions. The feature dimension for Local binary pattern is 257 x 493 features. Principle component analysis reduced the LBP feature's dataset, and it gives 257 x 1 dimensional data. To test the execution of the proposed system 75 images have been used for testing. The accuracy of the proposed system has been measure by comparing it with ground truth regions obtained by pre-processing, segmentation and ROI. Multi-SVM classifier gives the accuracy of 74.67%, FFNN classifier accomplishes the accuracy rate of 76 %, and DBN achieves the accuracy rate of 83% in terms of testing. The abnormality has been segmented into 4 grades. DBN gives the best accuracy rate for classifying the severity of osteoarthritis to their respective grades. In this work, two features, namely LBP and PCA, were computed. For classification purpose FFNN, Multi SVM and DBN are employed. When compared the classification result one with each other, the DBN gives the best accuracy than the other two works in .The proposed method has achieved the maximum classification accuracy of 83%. The performance comparison of the classifiers is shown in the tables. The graphical representation for classification accuracy is shown in the figures.

Table 2 Training and Testing dataset

Number of x-rays	126
Number of x-rays used for training	51
Number of x-rays used for testing	75

Table 3 Classification Metrics for the Multi-SVM

Class	Precision	Recall	F-measure
Grade-1	0.6875	0.7333	0.7097
Grade-2	0.6471	0.7181	0.6875
Grade-3	0.8462	0.7035	0.7857
Grade-4	0.8462	0.7323	0.7857
Mean	0.7567	0.7218	0.7421

Table 4 Classification Metrics for the ANN (FFNN)

Class	Precision	Recall	F-measure
Grade-1	0.6667	0.8	0.7273
Grade-2	0.8462	0.7333	0.7857
Grade-3	0.8	0.8	0.8
Grade-4	0.7857	0.7333	0.7586
Mean	0.7746	0.7666	0.7679

Table 5 Classification Metrics for the DBN

Class	Precision	Recall	F-measure
Grade-1	0.7222	0.8667	0.7879
Grade-2	0.801	0.78	0.795
Grade-3	0.9231	0.8	0.8571
Grade-4	0.9286	0.8667	0.8966
Mean	0.8437	0.8283	0.8341

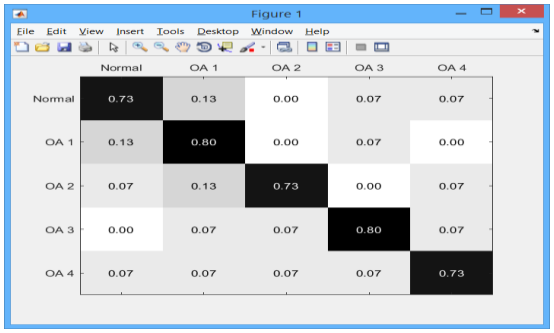


Fig. 13 (a) Confusion Matrix for FFNN

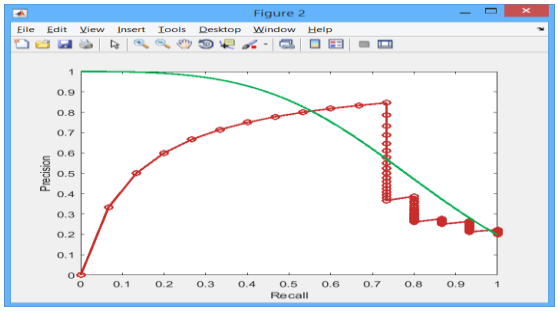


Fig. 13 (b) Confusion Matrix for FFNN

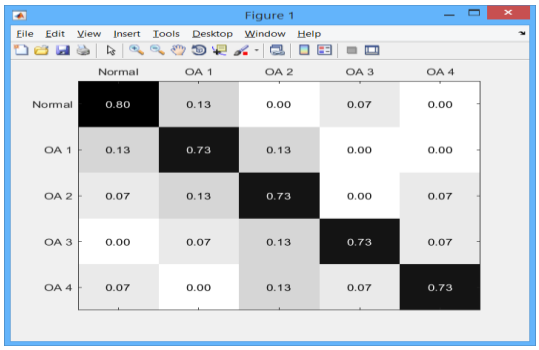


Fig. 14 (a) Confusion Matrix for SVM

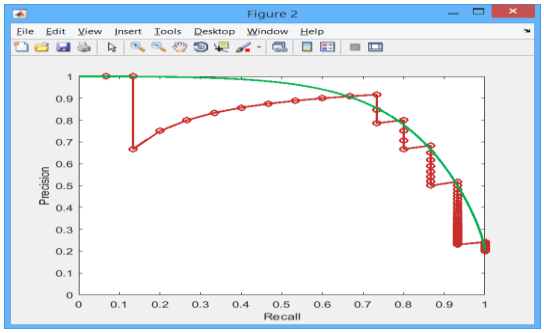


Fig. 14 (b) Confusion Matrix for SVM

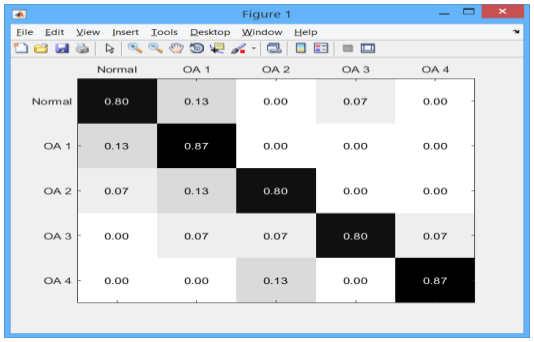


Fig. 15 (b) Confusion Matrix for DBM

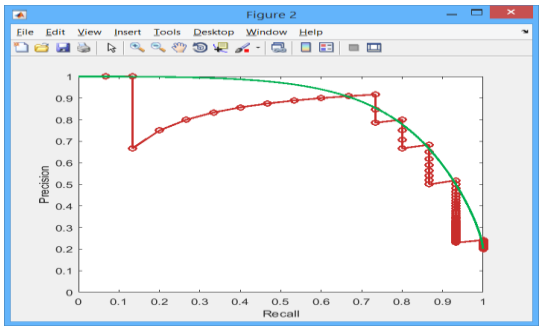


Fig. 15 (b) Confusion Matrix for DBM

The formula of evaluating metrics is given below:

$$Precision = \frac{TP}{TP} + FP$$

$$Recall = \frac{TP}{TP} + FN$$

$$F - Measure = 2.Precision.\frac{Recall}{Precision} + Recall$$

Conclusion

In this work, we have presented a system for diagnosing and classification of OA using machine learning methods. The proposed work has accomplished a reasonable accuracy rate of 83% for the classification of the disease. The self-assessment scoring framework may be valuable for recognizing the grown-up risk for knee OA. The execution of the framework has been improved by Multi-Svm, DBN, and FFNN. In the study, we have presented a computerize computer-aided determination approach for the discovery of OA, which makes utilisation of a combination of normalisation based on prescient modelling and feature extraction. Our analysis reveals that the proposed framework can give high classification execution in recognizing between healthy and OA affected knee. In terms of execution, it beats a few strategies proposed in the related work. We hypothesise that such an investigation approach might serve for the early discovery of OA. In this paper, a necessary, non-invasive, proficient and precise discovery for the early detection of knee OA has always been an interest within the medical field. The combination of computer innovation and the advanced plain

radiographs procedure helps in identifying the disease and disposes of the impact of the subjective variables in traditional morphological determination with OA pictures, hence avoiding the significant distinction in symptomatic outcomes caused by differences in information and experience. It can identify the existing risk of OA even before the morphological changes, and in this way, it can deliver more reliable outcomes of OA and treatment practical assessment.

At last, our examination affirms recent recommendations that radiographic reading in multicenter studies ought to be continuously performed centrally by an experienced investigator as the intra-observer unwavering quality of about all recorded features clearly depends on the individual involvement of the reader. In conclusion, our experiment has attempted to overcome a few distinguished restrictions of recent reliability studies about scoring frameworks. The result shows that a dependable radiographic classification of the seriousness of knee OA is possible by appropriate and joint location, particularly individual features as well as by overall evaluation. The limited reproducibility of certain features like (subchondral sclerosis, osteophytes, and cysts) needs further improvement in the classification process.

Our experiment reveals that more than 93% of moderate OA cases were separated precisely from ordinary cases with a false positive rate of 11.3%. The accuracy for the classification of separating minimal OA from the typical case was 83%. The future attempt for improving the accuracy for doubtful OA includes bodyweight, knee damage, history of knee, and will utilize more x-ray image as they got to be accessible. It is conceivable that a 100% relationship between computer-based and manual classification may not be accomplished. We recognize that the technique used to get the knee image does not permit for maximal determination of joint structures. We hypothesise that radiographic pictures would typically permit for the superior depiction of OA grades.

References

1. Lior Shamir et al., "Knee X-Ray Image Analysis Method for Automated Detection of Osteoarthritis", *IEEE Transaction on BIOMEDICAL Engineering*, vol. 56, No. 2, Feb 2009.
2. H. Oka, S. et al., "Fully automatic quantification of knee osteoarthritis severity on plain radiographs," *Osteoarthritis and Cartilage*, vol. 16, no. 11, 2008, pp. 1300– 1306.
3. L. Shamir, *et al*, "Early detection of radiographic knee osteoarthritis using computer-aided analysis," *Osteoarthritis and Cartilage*, vol. 17, no. 10, 2009, pp. 1307–1312.
4. H. J. Braun and G. E. Gold, "Diagnosis of osteoarthritis: Imaging," Bone, vol. 51, no. 2, 2012, pp. 278–288.
5. H. Oka, S. *et al*., "Fully automatic quantification of knee osteoarthritis severity on plain radiographs," *Osteoarthritis and Cartilage*, vol. 16, no. 11,, 2008, pp. 1300– 1306.
6. J. Kellgren and J. Lawrence, "Radiological assessment of osteoarthritis," *Annals of the rheumatic diseases,* vol. 16, 1957, pp. 494.
7. A. Guermazi, *et al*., "Prevalence of abnormalities in knees detected by MRI in adults without knee osteoarthritis: population-based observational study (Framingham Osteoarthritis Study)," *BMJ: British Medical Journal,* vol. 345, 2012.
8. W. Wirth, *et al*., "Direct Comparison of Fixed Flexion Radiography and MRI in Knee Osteoarthritis: Responsiveness Data from the Osteoarthritis Initiative," *Osteoarthritis and Cartilage,* 2012.
9. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthrosis. *Ann Rheum Dis*. 1957; 16 (4):494–502.
10. R. K. Arya, Vijay Jain, "Osteoarthritis of the knee joint: An overview", JIACM, 14(2), 2013, pp. 154-62.
11. Dileep Kumar, *et al*, "Development of a Non-invasive Diagnostic Tool for Early Detection of Knee Osteoarthritis", 978-1-4577-1884-7/11 IEEE, 2011.
12. R.S. Hegadi et al., Block based texture analysis approach for knee osteoarthritis identification using SVM. In: 2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), pp. 338– 341, 2015, IEEE.
13. L Anifah, et al., "Osteoarthritis classification using self-organizing map based on Gabor kernel and contrast-limited adaptive histogram equalization". Open Biomed. Eng. J. **7**, 18 (2013).
14. Brahim, A, *et al*., "A decision support tool for early detection of knee osteoarthritis using x-ray imaging and machine learning", Data from the osteoarthritis initiative. Computational. Med. Imaging Graph. 2019, 73, 11–18.
15. Shamir Lior, *et al*. "Early Detection of Radiographic Knee Osteoarthritis using Computer-Aided Analysis", *National Institutes of Health(NIH) Public Access, Osteoarthritis Cartilage*, Volume 17, Issue 10, October 2009.
16. Tati L et al, "Automated Detection of Unimpaired Joint Space for Knee Osteoarthritis Assessment" pp 400-403, IEEE 2005.
17. Aleksei Tiulpin, *et al*." A novel method for automatic localization of joint area on plain knee radiographs", A *Scandinavian Conference on Image Analysis*, 1 Feb 2017, DOI: 10.1007/978-3-3-319-59129-2-25.
18. Pooja P, *et al."* Severity analysis of Osteoarthritis of knee joint from X-ray images: A Literature review", 2014 *International Conference on Signal propagation and computer technology (ICSPCT 2014),* 2014 DOI:

10.1109/ICSPCT.2014.6885008, ISBN: 978-1-4799-3140-8, ISBN: 978-1-4799-3139-2, ©2005IEEE, pp.648-652.

19. Banerjee S, et.al. "Osteophyte Detection for Hand Osteoarthritis Identification in X-ray Images using CNNs", *33rd Annual International Conference of the IEEE EMBS Boston*, Massachusetts USA, September 3, 2011.
20. G.W. Stachowiak et al., "Detection and prediction of osteoarthritis in knee and hand joints based on the X-ray image analysis", *Bio surface and Biotribology*, Volume 2, Issue 4, December 2016, DOI: 10.1016/j.bsbt.2016.11.004, pp. 162-172.
21. X. Tan and B. Triggs, "Fusing Gabor and LBP feature sets for kernel-based face recognition," in Analysis and Modeling of Faces and Gestures. Springer, 2007, pp. 235–249.
22. T. Ahonen and M. Pietikainen, "A framework for analysing texture descriptors," Threshold, vol. 5, no. 9, 2008, pp. 1.
23. G. Zhenhua, *et al*., "A Completed Modeling of Local Binary Pattern Operator for Texture Classification," *Image Processing, IEEE Transactions on*, vol. 19, pp., 2010, 1657-1663.
24. Stojanovic, *et al*.," Impact of PCA based fingerprint compression on matching performance." In Telecommunications Forum (TELFOR), 2012 20th, pp. 693- 696. IEEE, 2012.
25. Beltran, Luis A. "Nonparametric multivariate statistical process control using principal component analysis and simplified depth." PhD diss., University of Central Florida, Orlando, Florida, 2006.
26. P.Dian, *et al*., "An application of back propagation Artificial Neural Network method for measuring the severity of OA", International journal of Engineering and Technology, Vol11, no. 3, June 2011, PP 102-105.
27. Laura Auria, Rouslan A. Moro, "Support Vector Machines (SVM) as a Technique for Solvency Analysis"
28. https://Wiki Svm, http://www.dataminingtools.net/ wiki/svm.php.
29. http://www.dtreg.com/svm.html
30. Data Mining Concepts and Techniques book By Jiawei Han and Micheline Kambe.
31. Chih-Wei Hsu and Chih-Jen Lin, "A Comparison of Methods for Multiclass Support Vector Machines", IEEE Transactions on Neural Networks, Vol. 13, No. 2, March 2002.
32. http://sci2s.ugr.es/ovo-ova/.
33. Le Roux N, Bengio Y. Representational power of restricted Boltzmann machines and deep belief networks [J]. Neural Computation, 2008, 20(6): 1631-1649.
34. G. E. Hinton, "A fast learning algorithm for deep belief nets", Neural Computation, vol. 18, pp. 15271554, 2006.
35. Le Roux N, Bengio Y. Representational power of restricted Boltzmann machines and deep belief networks [J]. Neural Computation, 2008, 20(6): 1631-1649.
36. Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann machines for collaborative filtering[C] Proceedings of the 24th international conference on Machine learning. ACM, 2007: 791-798.
37. Geoffrey Hinton, "A practical guide to training restricted Boltzmann machines," *Momentum*, vol. 9, no. 1, March 2002, pp. 926.