ISSN No:-2456-2165

Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity using Modified Fully connected Convolutional Neural Networks

[1]D.Kiruthika, [2]J. Judith
[1]Junior Research fellow, ECE, [2]Assistant professor, ECE,
Sethu Institute of Technology, Kariapatti, Virudhunagar, Tamilnadu, India

Abstract:- Knee Osteoarthritis (OA) is an extremely common and degenerative musculoskeletal disease worldwide which creates a significant burden on patients with reduced quality of life and also on society because of its financial impact. Therefore, technical try and efforts to reduce the burden of the disease could help both patients and society. In this paper, an automated novel method is proposed with a supported combination of joint shape and modified Fully connected neural network (FCNN) based bone texture features, to differentiate between the knee radiographs with and without osteoarthritis. Moreover, an endeavor is formed to explain the bone texture using CNN. Knee radiographs from Osteoarthritis Initiative (OAI) and Multicenter Osteoarthritis (MOST) datasets are utilized in this paper. The proposed models were trained on 8000 knee radiographs from OAI and evaluated on 3500 knee radiographs from MOST. The results demonstrate that fusing the proposed shape and texture parameters achieves the state-of-the art performance in radiographic OA detection yielding area under the ROC curve (AUC) of 98.75% accuracy.

Keywords:- *Knee Osteoarthritis, KL grades, Automatic Detection, Fully Convolutional Neural Networks, Classification and Regression.*

I. INTRODUCTION

Knee Osteoarthritis (OA) may be a debilitating joint disorder that mainly degrades the knee articular cartilage. Clinically, the most important pathological features of knee OA include joint space narrowing, osteophyte formation, and sclerosis. Knee OA has a high incidence among the elderly, obese, and people with a sedentary lifestyle. In its severe stages, it causes excruciating pain and ends up in total joint arthroplasty. In the Early diagnosis period , it is a crucial for clinical treatments and pathology [1]. Despite the introduction of several imaging modalities like, Optical Coherence Tomography,MRI and radiography (X-ray) , an ultrasound for augmented OA diagnosis, has been traditionally preferredand “gold standard” for preliminary diagnosis of knee OA [2]. During this work, we train CNNs from scratch to automatically quantify knee OA severity using X-ray images. This involves two main steps: 1) automatically detecting and extracting the region of interest (ROI) and localizing the knee joints, and 2) classifying the localized knee joints.

To automatically localise the knee joints, we present a fully-convolutional neural network (FCN)-based method. A FCN is an end-to-end network that has been trained to make pixel-by-pixel predictions [3]. To automatically classify the localized knee joints, we propose two methods: 1) training a CNN from scratch for multi-class classification of knee OA images, and 2) training a CNN to optimize a weighted ratio of two loss functions: categorical cross-entropy for multi-class classification and mean-squared error for reclassification. We compare the results of these methods to those of WND-CHARM [4] and our previous research. We also compare the classification results to both manual and automated methods.We propose a novel pipeline for automatically quantifying knee OA severity, which includes an FCN for localising knee joints and a CNN jointly trained for knee joint classification and regression. The main contributions of this work include a fully-convolutional network (FCN)-based method for automatically localising knee joints, as well as training a network (CNN) from scratch that optimises a weighted ratio of both categorical cross-entropy and mean-squared error for knee joint regression. This multi-objective convolutional learning improves overall quantification while also providing multi-class classification and regression outputs at the same time.

II. LITERATURE SURVEY

Classifying the severity of knee OA can be accomplished by detecting variations in joint space width and the formation of osteophytes in the knee joints [5]. Yoo et al., in a recent approach, developed a scoring system to predict radiographic and symptomatic knee OA risks using artificial neural networks (ANN) and KNHANES V-1 data. Then, the Shamir et al., proposed a WND-CHARM method , which is a multipurpose bio-medical image classifier [6] was used to classify knee OA radiographs [7] and to detect knee OA early using computer aided analysis. WND-CHARM extracts handcrafted features from raw images and image transforms [8].

Convolutional neural networks (CNNs) have recently outperformed many methods based on hand-crafted features in many computer vision tasks such as image recognition, automatic detection and segmentation, content-based image retrieval, and video classification. CNNs learn effective feature representations that are particularly well-suited for fine-grained classification [9], such as knee OA image classification. In a previous study, we demonstrated that off-the-shelf CNNs trained on the ImageNet LSVRC dataset [13], such as the VGG 16-Layers network [10], the VGG-M-128 network [11], and the BVLC reference CaffeNet [12], can be fine-tuned for classifying knee OA images using transfer learning. We also argued that instead of binary or multi-class classification, it is preferable to assess knee OA severity using a continuous metric such as mean-squared error and showed that predicting the continuous grades through regression reduces the mean-squared error and in turn improves the overall quantification. Before that, Shamir et al. et al. [14] proposed using template matching to detect and extract knee joints automatically. For large datasets like OAI, this method is slow, and the accuracy and precision of detecting knee joints is low. In a previous study, we presented an SVM-based method for automatically detecting the centre of knee joints [15] and extracting a fixed region as the ROI with reference to the detected centre. This method is also not very accurate, and the aspect ratio of the extracted knee joints is compromised, which affects the overall quantification.

III. MATERIALS AND METHODS

A. Data Acquisition

The data used in this study for the experiments and analysis are bilateral PA fixed flexion knee X-ray images. The datasets are from the University of California, San Francisco's Osteoarthritis Initiative (OAI) and Multicenter Osteoarthritis Study (MOST), and are standard datasets used in knee osteoarthritis studies.

B. Kellgren and Lawrence Grades

The Kellgren and Lawrence (KL) grades are used as the ground truth in this study to classify the knee OA X-ray images. The KL grading system is still regarded as the gold standard for assessing the severity of knee osteoarthritis in radiographs [16]. It assigns a five-point scale to the severity of radiographic knee OA. 'Grade 0' means normal, 'Grade 1' means doubtful, 'Grade 2' means minimal, 'Grade 3' means moderate, and 'Grade 4' means severe. The KL grading system is depicted in Figure 1.

C. OAI and MOST Datasets

The OAI dataset's baseline cohort includes MRI and X-ray images from 4,476 participants. We chose 4,446 X-ray images from the entire cohort based on the availability of KL grades for both knees as determined by the Boston University X-ray reading centre (BU). There are 8,892 knee images in total, with the following distribution based on KL grades: Grade 0 - 3433, Grade 1 - 1589, Grade 2 - 2353, Grade 3 - 1222, and Grade 4 - 295.The MOST dataset contains lateral knee radiograph assessments from 3,026 people. Based on the availability of KL grades for both knees as per baseline to 84-month Longitudinal Knee Radiograph Assessments, 2,920 radiographs are chosen. This dataset contains 5,840 knee images and the distribution as per KL grades is as follows: Grade 0 - 2497, Grade 1 - 1017, Grade 2 - 922, Grade 3 - 970, and Grade 4 - 431.

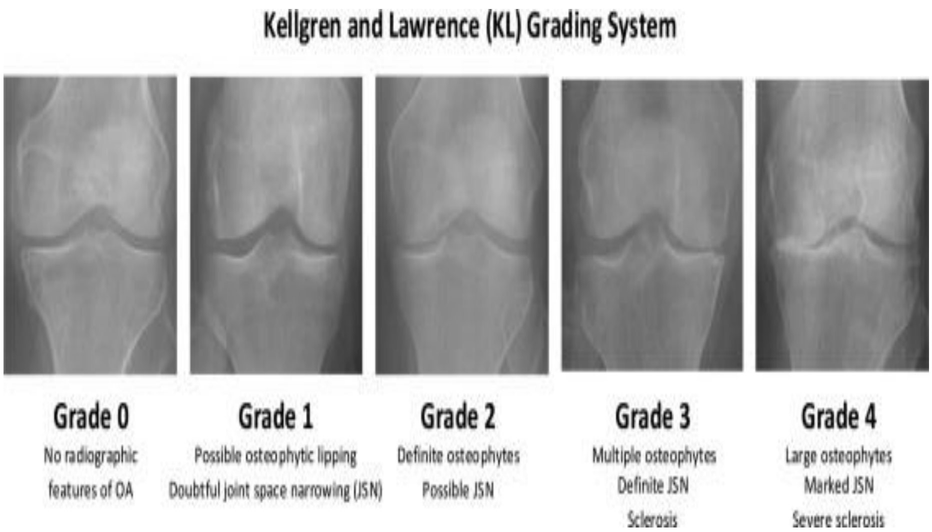


Fig. 1: The KL grading system to assess the severity of knee OA [16]

IV. PROPOSED METHODOLOGY

This section describes the methodology for calculating radiographic knee OA severity. This consists of two steps: automatically detecting knee joints with a fully convolutional network (FCN) and simultaneously classifying and localized knee images with a convolutional neural network (CNN). Figure 2 depicts the entire pipeline used to assess the severity of knee OA.

A. Automatically Localizing Knee Joints using a FCN architecture

The severity of knee OA can be determined by detecting variations in joint space width and the formation of osteophytes in the knee joint [17]. Thus, localizing the knee joints from X-ray images is an important pre-processing step before quantifying knee OA severity, and automatic methods are preferable for larger datasets. Figure 3 depicts a knee OA radiograph and the region of interest (ROI) for detection. Previous methods for automatically localizing knee joints, such as template matching [18] and our own SVM-based method, were ineffective. In this paper, we propose a fully convolutional neural network (FCN)-based method for improving the accuracy and precision of detecting knee joints.

B. Fully connected Convolutional Neural Network Architecture

We trained a fully convolutional neural network (FCN) to automatically detect the region of interest (ROI): the knee joints from knee OA radiographs, inspired by the success of a fully convolutional neural network (FCN) for semantic segmentation on general images. Our proposed FCN is built on a lightweight architecture, and the network parameters are learned from the scratch. The architecture is depicted in Figure 2. After the experiment analysis, we discovered that this architecture is the best for detecting knee joints.

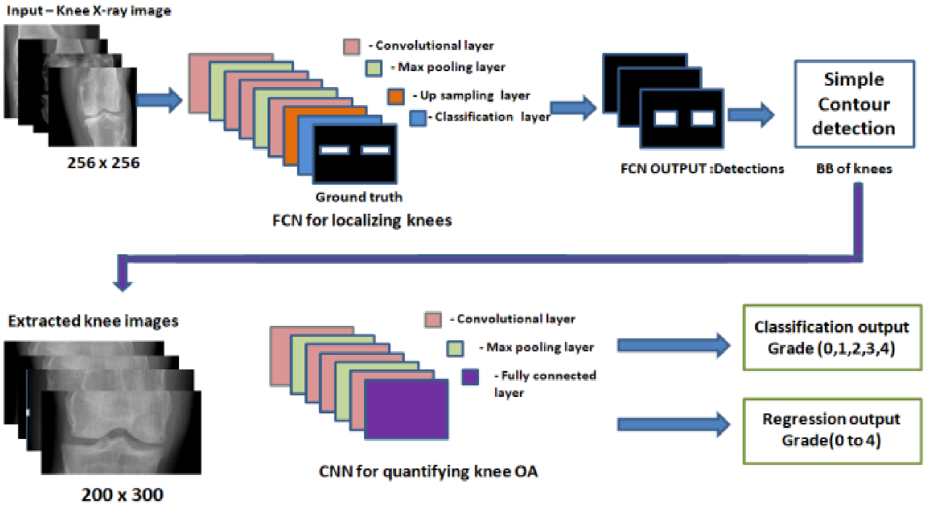


Fig. 2: The proposed architecture for detecting Knee Osteoarthritis.

The network is composed of four stages of convolutions, each followed by a max-pooling layer, and the final stage of convolutions is followed by an up-sampling and a fully-convolutional layer. The first and second layer of convolution employ 32 filters each, the third layer consist of 64 filters, and the fourth layer with 96 filters. The network employs uniform [3 3] convolution and max pooling [2 2]. Following each convolution layer is a batch normalization and a rectified linear unit activation layer (ReLU). As the network employs three stages of [2 2] max pooling, a [8 8] up-sampling is performed after the final convolution layer. Up-sampling is required for end-to-end learning via back propagation from pixel-wise loss and to obtain pixel-dense results. For pixel-based classification, the final layer is a fully convolutional layer with a kernel size of [1 1] and sigmoid activation. The network's input is [256 256], and its output is the same size.

C. FCN Training

The network was trained from scratch using training samples of knee OA radiographs from the OAI and MOST datasets. Binary images with masks specifying the ROI: the knee joints are used to train the network. Figure 4 depicts an example of a binary mask: the ground truth. The binary masks were generated from manual annotations of knee OA radiographs using a fast annotation tool that we created. The network was trained to minimise the total binary cross entropy between predicted and true pixels. We used the adaptive moment estimation (Adam) optimizer [20] with default parameters and discovered that it provided faster convergence than standard SGD. Figure 3 shows that the ROI (Region of Interest) regions are marked and segmented.

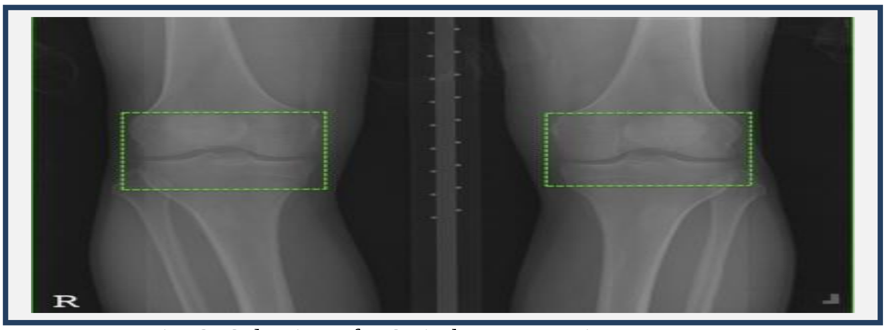


Fig. 3: Selection of ROI in knee X-ray image

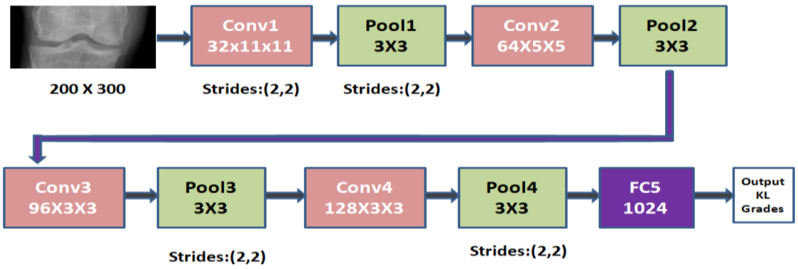


Fig. 4: The Fully Convolutional Network for automatically detecting knee joints

In the figure 4, the knee joints are extracted and deduce the bounding boxes of the knee joints using simple contour detection from the output predictions of FCN. n the proposed architecture, four convolutional layers are added , then four Max-pooling layers are added to reduce the dimensionality of the datasets. The fully connected layers are added at the end of the architecture. The output of the max-pooling layers is given as an input to the fully connected layers. The output from this layers are flattened, the weights are compared with the flattened value to predict the Normal and OA images, and extract the knee joints from knee OA radiographs using the bounding boxes. We upscale the bounding boxes from the FCN output of size [256 256] to the original size of each knee OA radiograph before extracting the knee joints to preserve the aspect ratio of the knee joints.

D. Quantifying knee OA severity using Convolutional Neural Network

Convolutional neural networks are trained from scratch using knee OA data, and networks are jointly trained to minimize classification and regression losses, allowing for a more accurate assessment of knee OA severity.

- **Training CNN for Classification:** The network is made up of five learned weight layers: four convolutional layers and one fully connected layer. The network architecture is depicted in Figure 5. Due to the scarcity of training data, we considered a lightweight architecture with few layers, and the network has 5.4 million free parameters in total. We discovered that this architecture is the best for classifying knee images after experimenting with the number of convolutional layers and other parameters. The network's convolutional layers are followed by batch normalization and a rectified linear unit activation layer (ReLU). There is a max pooling layer after each convolutional stage. Following the final pooling layer is a fully connected layer and a softmax dense layer. We include a drop out layer to avoid over-fitting. In the final two convolutional layers (conv3 and conv4), as well as the fully connected layer, we apply an L2-norm weight regularization penalty of 0.01. (fc5). Applying a regularization penalty to other layers lengthens training time while introducing no significant variation in learning curves. Using the Adam optimizer, the network was trained to minimize categorical cross-entropy loss [20]. The network's inputs are knee images with dimensions of [200 300]. Based on the mean aspect ratio (1.6) of all extracted knee joints, we chose this size to roughly preserve the aspect ratio.

E. Jointly training CNN for Classification and Regression

In general, assessing the severity of knee OA is based on multi-classifying knee images and assigning a KL grade to each distinct category [10]. We argued in our previous paper [1] that assigning a continuous grade (0-4) to knee images through regression is a better approach for quantifying knee OA severity because the disease is progressive. However, there is no ground truth of KL grades on a continuous scale to train a network directly for regression output with this approach. As a result, we use multi-objective convolutional learning [8] to train networks to optimize a weighted-ratio of two loss functions: categorical cross-entropy and mean-squared error. Mean squared error informs the network about grade ordering, while cross entropy informs it about grade quantization. Optimizing a network with two loss functions intuitively provides a stronger error signal and is a step toward improving overall quantification, taking both classification and regression results into account. We arrived at the final architecture shown in Figure6 after much experimentation. This network has six learned weight layers: five convolutional layers and one fully connected layer, with a total of approximately 4 million free parameters. Following each convolutional layer is batch normalization and a rectified linear activation (ReLU) layer. We include drop out (p = 0.5) in the fully connected layer (fc5) and L2 weight regularization in the fully connected layer (fc5) and the final stage of convolution layers to avoid over-fitting this model (Conv3-1 and Conv3-2). We used stochastic gradient descent with Nesterov to train the model. The initial learning rate was set to 0.001, and it was reduced by a factor of 10 if the validation loss did not decrease for four consecutive epochs.

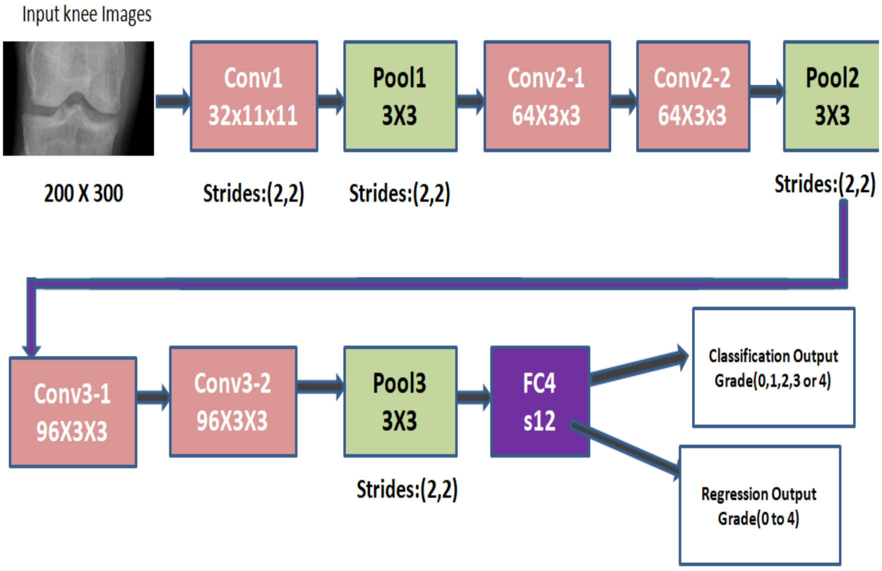


Fig. 5: The proposed network architecture for simultaneous classification and regression

V. RESULTS AND DISCUSSION

The proposed method trained FCNs to automatically localize and extract knee joints from X-ray images of knee OA. The datasets are divided into two parts: training/validation (70%) and test (30%). The OAI dataset contains 3,146 images for training and 1,300 images for testing. MOST dataset training and test samples are 2,020 and 900 images, respectively. First, we trained the network with OAI dataset training samples before testing it with OAI and MOST datasets separately. Following that, we expanded our training samples by including the MOST training set, and the test set is a hybrid of the OAI and MOST test sets. We experimented with the number of convolution stages, the number of filters, and the number of layers before settling on the final architecture. The final network (shown in Figure 4) was trained using samples from both the OAI and the MOST datasets.

A. Classification of Knee OA Images using a CNN

We use the same train-test split for localization and quantification to maintain pipeline consistency and allow valid comparisons of results obtained using different approaches. To increase the training samples, we include the right-left flip of each knee joint image, which doubles the total number of training samples available. As a first step, we trained neural networks to classify manually annotated knee joint images. After much experimentation, we arrived at the final architecture depicted in Figure 5.We compare our network's classification results to the previous best results for automatically quantifying knee OA severity, WND-CHARM, a multipurpose medical image classifier [11]. The multi-class classification accuracy and mean-squared error of our network and WND-CHARM are shown in Table 2. The results show that our network was trained from the scratch for classifying knee OA images clearly outperforms WND-CHARM. Also these results show an These results also show an improvement over our previous reported methods [1], which used off-the-shelf networks like VGG nets and the BVLC Reference CaffeNet for classifying knee OA X-ray images using transfer learning. These gains are due to our network's lightweight architecture, which was trained from scratch with fewer (5.4 million) free parameters than BVLC CaffeNet, which had 62 million free parameters for the same amount of training data. The off-the-shelf networks were trained on a large dataset like ImageNet, which contains millions of images, whereas our dataset contains far fewer training samples (10, 000). In the following section, we show further improvements in the results for quantifying knee OA severity.

Network	Sensitivity	Specificity	Accuracy	F1-Score
WIND-CHARM method	89.6	79.78	86	90.63
Fractal dimension	92.9	87.88	91.16	93.81
Fine-Tuned BVLC Caffe Net	94.39	88.96	93.85	95.42
Proposed Model	99.651	99.4242	98.755	99.3

Table 1: Comparison of the proposed model with existing system

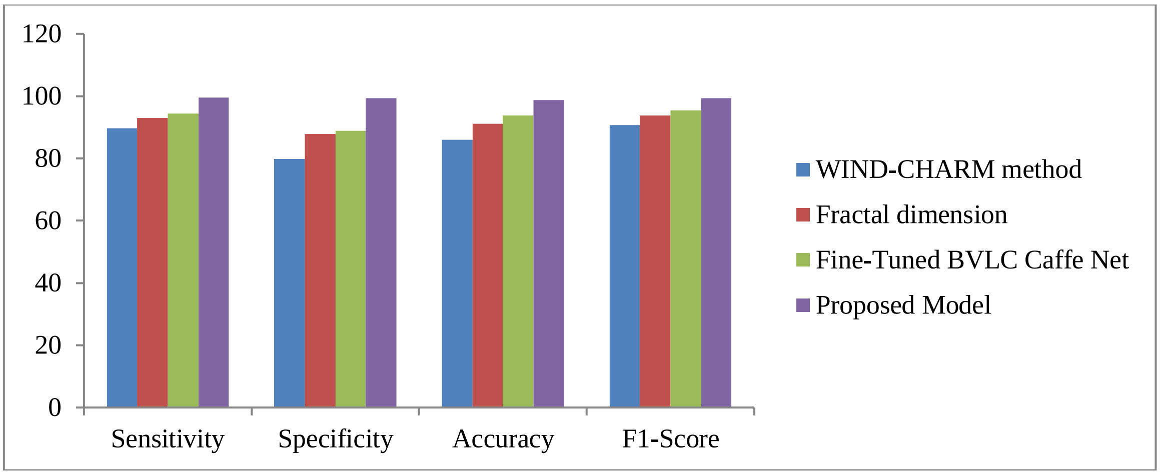


Fig. 6: Comparison of proposed work Results with the state of art methods

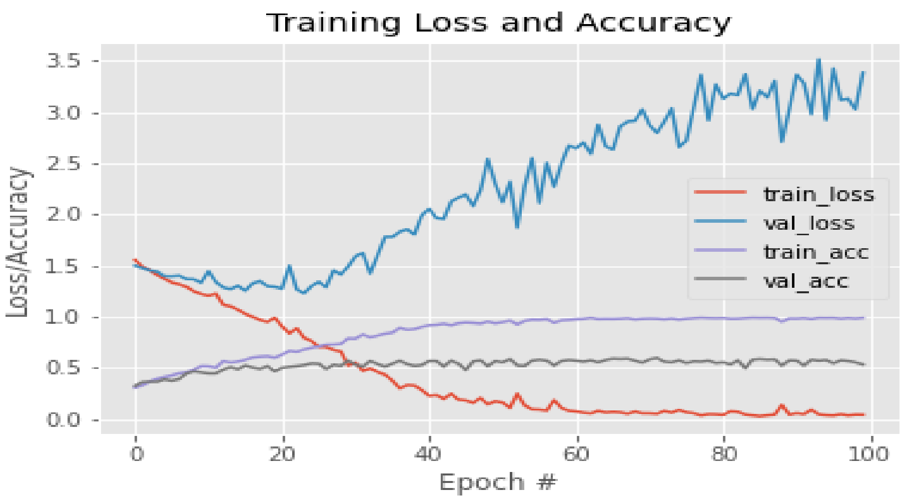


Fig. 7: (a) Training (Tr) and validation (Val) accuracy (acc), (b) Training and validation loss for joint classification(Clsf) and regression(Reg) training.

According to the results in Tables 2 and 3, the network trained jointly for classification and regression has a higher multi-class classification accuracy of 98.75% and a lower mean-squared error of 0.661 than the previous network trained only for classification, which has a multi-class classification accuracy of 98.75% and a mean-squared error of 0.898. Table 5 compares the precision, recall, F1 score, and area under curve (AUC) of the networks trained for classification and regression. These findings indicate that the network jointly trained for classification and regression learns a better representation than the previous network trained only for classification.

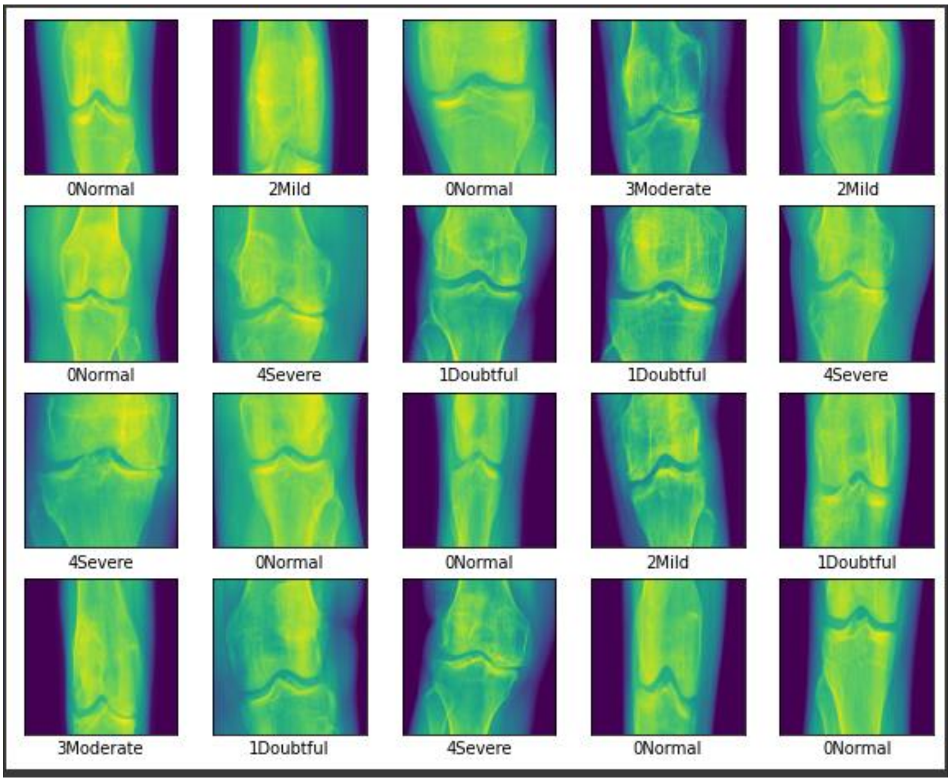
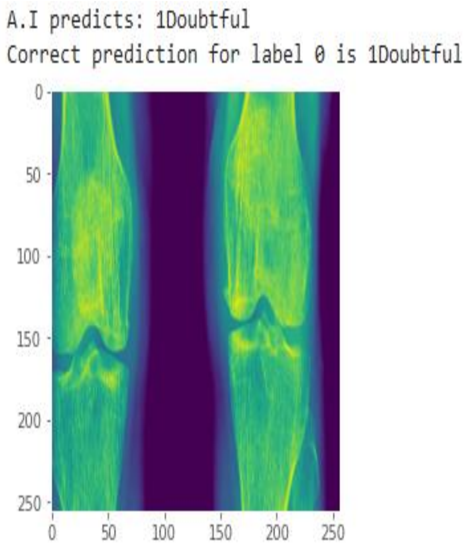


Fig. 8: Various classification of Osteoarthritis using OAI dataset

From the Figure 10 shows some examples of various classifications: knee joints predicted as Normal, Doubtful, Moderate, Mild and Severe. These images show minimal variations in terms of joint space width and osteophytes formation, making them challenging to distinguish. Even though the KL grades are used in clinical settings to assess knee OA severity, there has been ongoing investigation and criticism of their use because the individual categories are not equidistant from each other [3,4]. This could explain why the automatic quantification has a low multi-class classification accuracy. Because knee OA features such as joint space narrowing, osteophytes formation, and sclerosis are separately graded, using OARSI readings instead of KL grades may provide better results for automatic quantification.



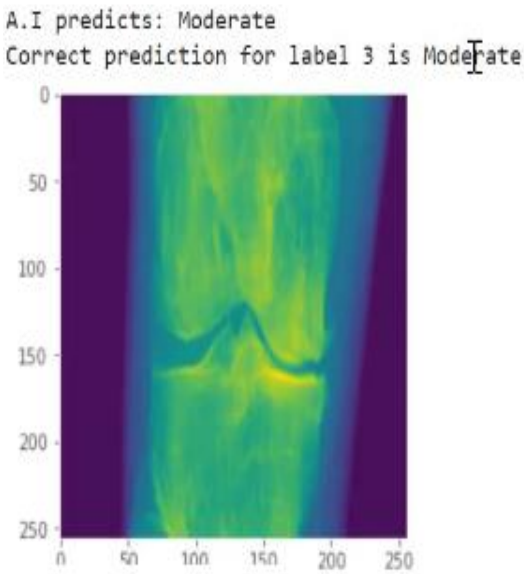


Fig. 9: Proposed model predicted the given input image as a) Doubtful b) Moderate

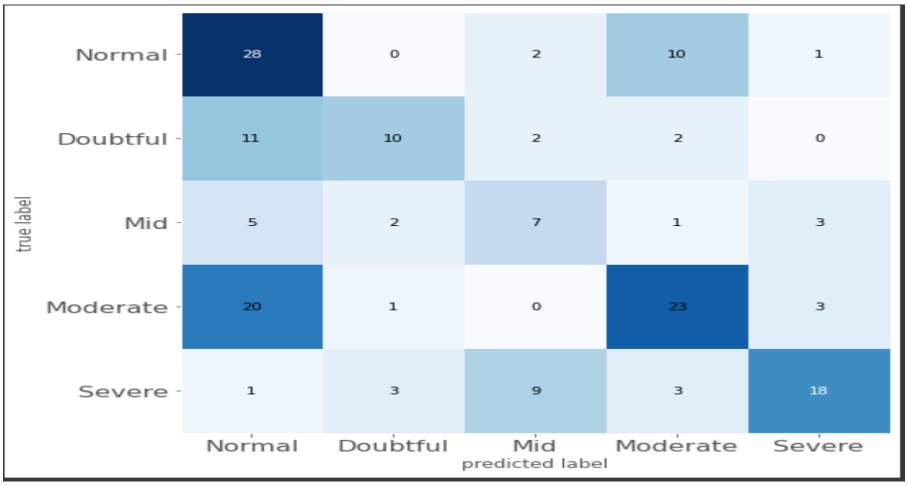


Fig. 10: The obtained Confusion matrix for the proposed model

VI. CONCLUSION

We proposed new methods for automatically localizing knee joints using a fully convolutional network and quantifying knee OA severity using a network jointly trained for multi-class classification and regression, both from scratch. In comparison to the previous methods, the FCN-based method is highly accurate. We demonstrated that the classification results obtained with automatically localized knee joints are comparable to the results obtained with manually segmented knee joints. In comparison to the previous method, the jointly trained network for classification and regression improves in multi-class classification accuracy, precision, recall, and F1 score. The confusion matrix and other metrics show that classifying Knee OA images conditioned on KL grade 1 is difficult due to small variations, especially in the consecutive grades from grade 0 to grade 2. Future work will focus on training an end-to-end network to quantify knee OA severity by integrating the FCN for localization and the CNN for classification. It will be interesting to compare the human-level accuracy involved in assessing knee OA severity to automatic quantification methods. This could provide insights into how to improve fine-grained classification even further.

REFERENCES

[1.] Antony, J., McGuinness, K., Connor, N.E., Moran, K.: Quantifying radiographic knee osteoarthritis ever it using deep convolutional neural networks .In: Proceedings of the 23rd International Conference on Pattern Recognition. IEEE(2016),InPress.

[2.] Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: Delving deep into convolutional nets. In: Proceedings of British Machine Vision Conference(2014).

[3.] Emrani, P.S., Katz, J.N., Kessler, C.L., Reichmann, W.M., Wright, E.A., McAlindon,T.E., Losina, E.: Joint space narrowing and Kellgren–Lawrence progression in knee osteoarthritis: an analytic literature synthesis. Osteoarthritis and Cartilage 16(8),873–882(2008).

[4.] Hart,D.,Spector,T.: Kellgren & Lawrence grade1 osteophytes in the knee doubtful or definite? Osteoarthritis and cartilage,11(2),149–150(2003).

[5.] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama,S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In:Proceedings of the ACM International Conference on Multimedia. pp. 675–678(2014)

[6.] Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A.,Winnemoeller,H.:Recognizingimagestyle.arXivpre printarXiv:1311.3715(2013)

[7.] Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprintarXiv:1412.6980(2014)

[8.] Liu, S., Yang, J., Huang, C., Yang, M.H.: Multi-objective convolutional learningfor face labeling. In: Proceedings of the IEEE Conference on Computer Vision andPatternRecognition.pp.3451–3459(2015)

[9.] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semanticsegmentation.In:ProceedingsoftheIEEEConferenceonComputerVisionandPatternRecognition.pp.3431–3440(2015)

[10.] Oka, H., Muraki, S., Akune, T., Mabuchi, A., Suzuki, T., Yoshida, H., Yamamoto,S., Nakamura, K., Yoshimura, N., Kawaguchi, H.: Fully automatic quantificationof knee osteoarthritis severity on plain radiographs. Osteoarthritis and Cartilage16(11),1300–1306(2008)

[11.] Orlov, N., Shamir, L., Macura, T., Johnston, J., Eckley, D.M., Goldberg, I.G.: WND-CHARM: Multi-purpose image classification using compound image transforms.Patternrecognitionletters29(11),1684–1693(2008)

[12.] Park,H.J., Kim, S.S., Lee, S.Y., Park, N.H., Park, J.Y., Choi, Y.J., Jeon, H.J.:A practical MRI grading system for osteoarthritis of the knee: association withKellgren–Lawrence radiographic scores. European journal of radiology 82(1), 112–117(2013)

[13.] Russakovsky,O.,Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang,Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: ImageNet large scale visualrecognition challenge. International

Journal of Computer Vision 115(3), 211–252(2015)

[14.] Shamir,L.,Ling,S.M., Scott, W., Hochberg, M., Ferrucci, L., Goldberg, I.G.:Early detection of radiographic knee osteoarthritis using computer-aided analysis.OsteoarthritisandCartilage17(10),1307–1312(2009)

[15.] Shamir, L., Ling, S.M., Scott Jr, W.W., Bos, A., Orlov, N., Macura, T.J., Eckley,D.M.,Ferrucci,L.,Goldberg,I.G.:KneeX-rayimageanalysismethodforautomateddetectionofosteoarthritis.IEEETransactionsonBiomedicalEngineering 56(2),407–415(2009)

[16.] Shamir, L., Orlov, N., Eckley, D.M., Macura, T., Johnston, J., Goldberg, I.: Wnd-charm:Multi-purposeimageclassifier.AstrophysicsSourceCodeLibrary(2013)

[17.] Shamir, L., Orlov, N., Eckley, D.M., Macura, T., Johnston, J., Goldberg, I.G.:Wndchrm–an open source utility for biological image analysis. Source code forbiologyandmedicine3(1),13(2008)

[18.] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scaleimagerecognition.arXivpreprintarXiv:1409.1556 (2014)

[19.] Yang, S.: Feature Engineering in Fine-Grained Image Classification. PhD Thesis,UniversityofWashington(2013)

[20.] Yoo, T.K., Kim, D.W., Choi, S.B., Park, J.S.: Simple scoring system and artificialneuralnetworkforkneeosteoarthritisriskprediction:Across-sectionalstudy.PloSone11(2),e0148724(2016