Software Proposal Document For Automatic Recognition of Musculoskeletal Disorders from Radiographs

Sara Awwad[1], Menna Mohamed[1], Farah Hisham[1], Amira Galal[1], Taraggy M Ghanim[1] and Ayman Nabil[1]
[1]Faculty of Computer Science,Misr International University

September 26, 2018

Abstract

Automatic Recognition of Musculoskeletal Disorders is one of the important applications of computers in clinical medicine. Radiographs must be classified as healthy or abnormal. A critical need for fast and inexpensive but reliable automatic classification systems should be satisfied. It is desired to discriminate between the various types of musculoskeletal disorders. Different approaches have been proposed recently in this field of pattern-recognition. Our proposed paper introduce a detailed survey on up-to-date published papers specialized in this point of research. Also a comprehensive summary of all the available data-sets is included. Finally, a comparison between the latest proposed approaches is studied.

1 Introduction

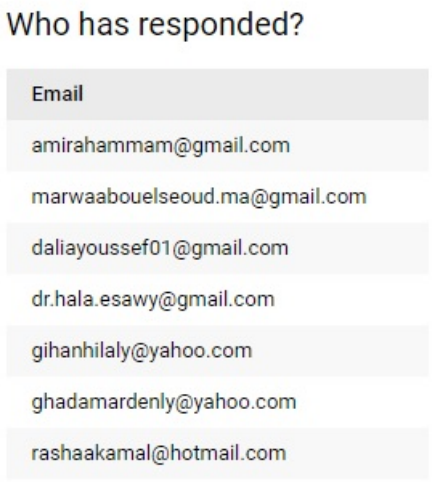
1.1 Background

Musculoskeletal disorders (MSDs) are becoming one of the most important health concerns. MSDs affect more than 1.7 billion people worldwide [1], and treatment is quite expensive and misclassification may cause severe side effects. MSDs are conditions that affect bones, joints and muscles. The different types of MSDs include tendinitis, carpal tunnel syndrome, osteoarthritis, rheumatoid arthritis and fibromyalgia and bone fractures [2]. Musculoskeletal imaging through radiographs involves acquisition, analysis and decision making. This can be applied on different musculoskeletal organs like bones, joints or even soft tissues due to injury. The automatic diagnosis of disorders serves treatment and helps in planning therapy and support interventional guidance.Recently the need of advanced computational methods that help in the diagnosis of MSDs has been increasing. Given a system that can automatically detect whether a radiograph is normal or abnormal based on musculoskeletal imaging radiographs, it would

help workflow and also the patients to be treated faster and accurately. It can also help reduce the workload on radiologists, which would mean less fatigue and thus more accurate. Automatic Recognition systems can help clinicians and researchers to introduce the commonness and occurrence rates, also it serves in determining the impact of disorders and their prognosis.

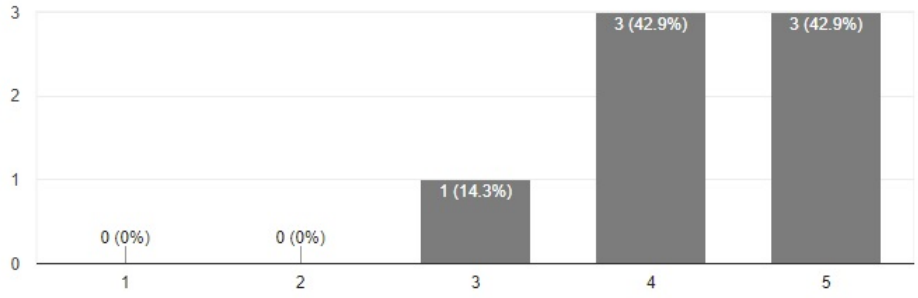
1.2 Motivation

There hasn't been any work that tackles the issues found in the elbow concerning MSD's. The only most recent work that developed an automatic recognition system was carried out by Stanford ([1]) called MURA, where they detect the existence of abnormalities in the upper-limbs from radiographs, and also have gathered the largest dataset containing 40,561 radiographs. However, merely detecting the existence of abnormalities is insufficient, and the reached automatic detection system is still not equivalent to the best radiologists. Since our focus is on the elbow, the dataset for the elbow from MURA's makes only a total of 1,912 images. That is why we will be working on enhancing the accuracy level of detecting abnormalities in the elbow and also classify the disorder present in the radiograph, and work on labelling the dataset. The results of the two surveys are as follows Doctors Survey:

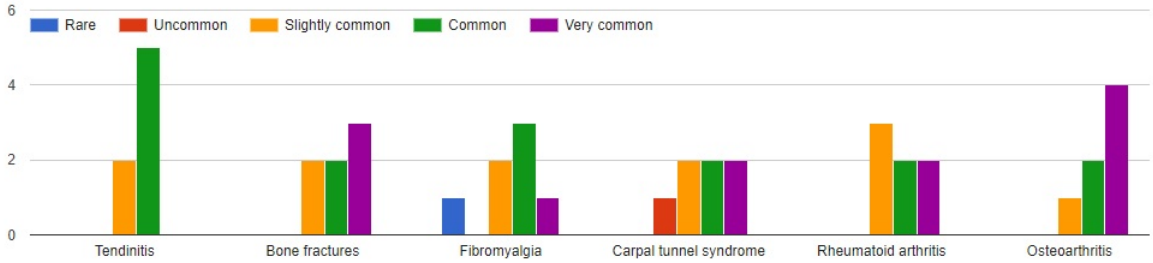


How common are MSDs?

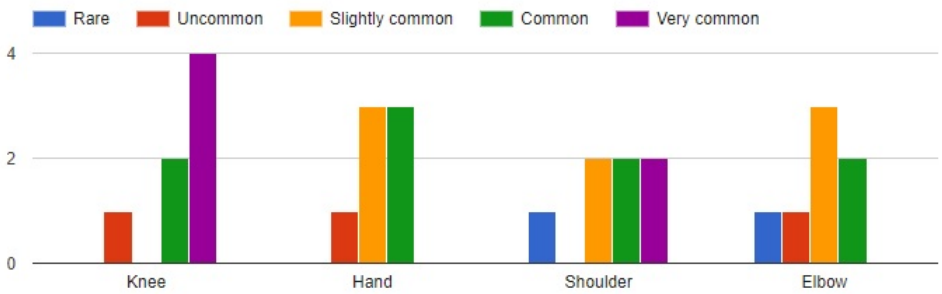
7 responses



How common are the following types of MSD's?

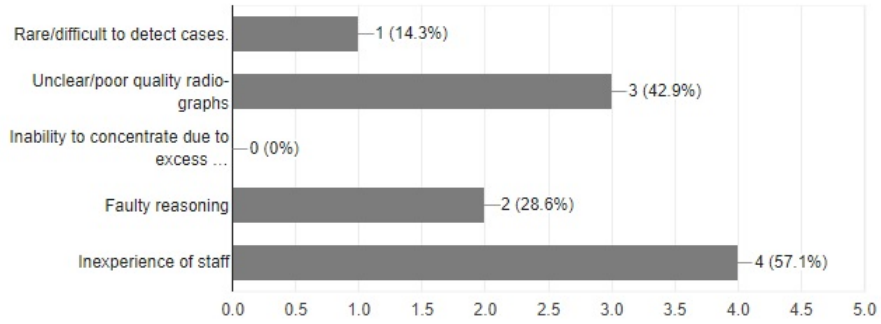


How common is it to suffer from MSD's in the following organs?



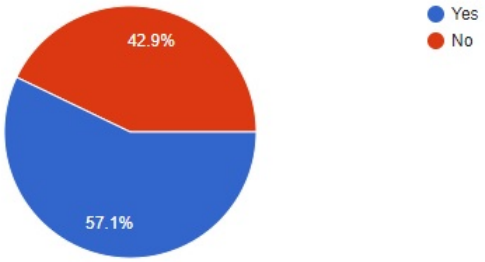
What are the problems radiologists face when diagnosing diseases from radio-graphs that can lead to misdiagnosis?

7 responses



Are there any features in a radio-graph that could be difficult to notice?

7 responses



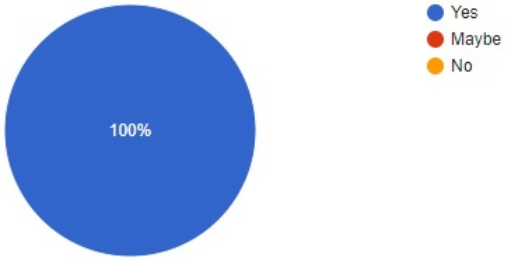
If yes, what are they?

3 responses

tendinitis
Early mcp erosions
Jount affection

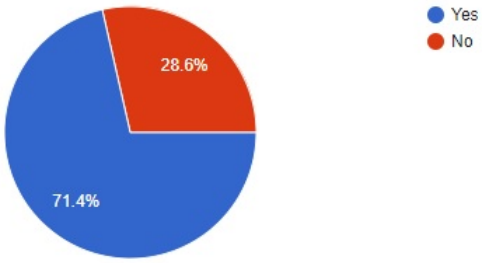
Does experience affect diagnosing a case correctly?

7 responses



Would you use a computer-aided system that would help classify the diagnosis of disorders from radiographs?

7 responses



Have you ever used/encountered a computer-aided program that would help with classifying a disorder, or do you have any suggestions that would help?

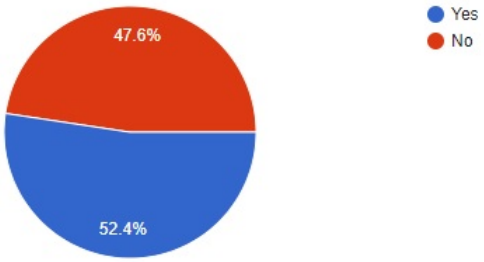
4 responses

No (3)
No i haven't

Patients Survey:

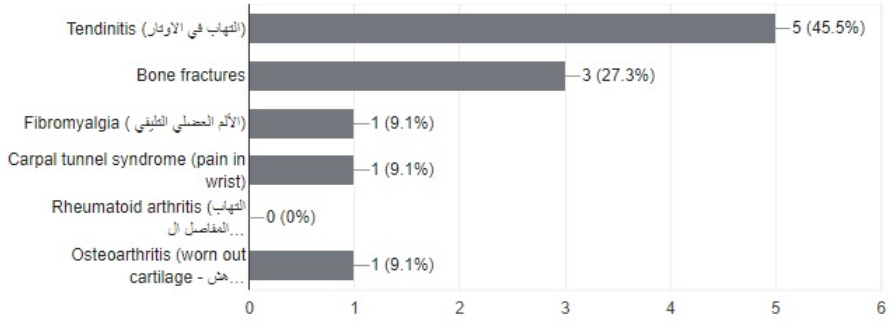
Have you ever had any disorders in your bones/joints/muscles?

21 responses



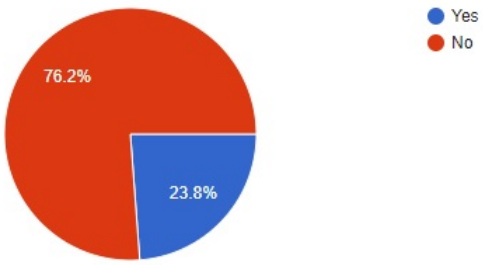
If yes, what type of disorder did you have?

11 responses



Have you ever been misdiagnosed by a radiologist or know someone who has?

21 responses



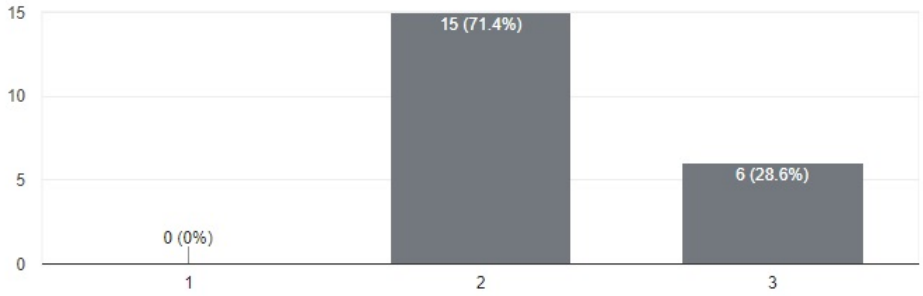
If yes, how did the misdiagnosis affect you/them?

5 responses

Increasing the disease
taking wrong medicine without getting better
i was not getting any better
Incorrect medication, extra money has been paid, and medicines' disadvantages occurred for nothing.
Had to take unnecessary antibiotic course

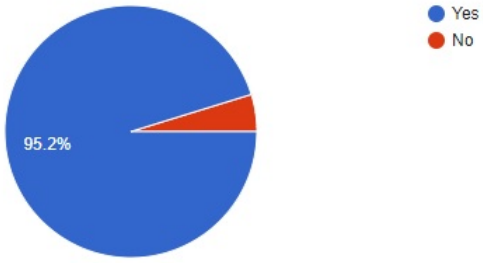
How much does a doctor visit cost you?

21 responses



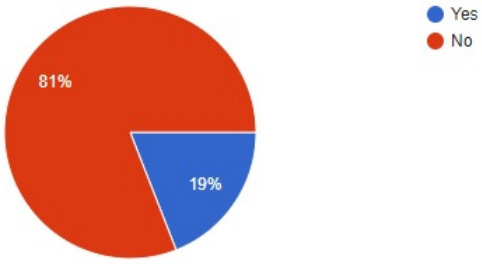
Do you think more experienced doctors are more expensive?

21 responses



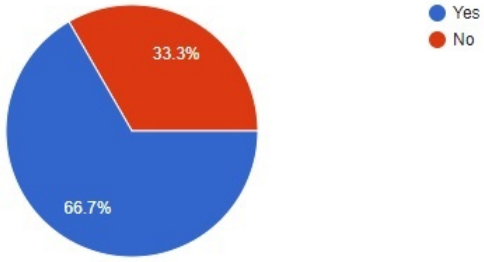
Is the appointment scheduling process smooth and fast for an experienced doctor?

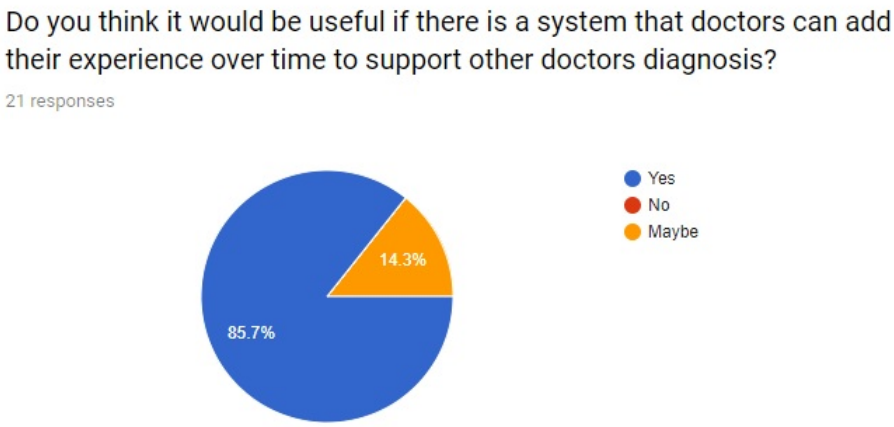
21 responses



Have you ever had to visit another doctor to proof-read another doctor's diagnosis?

21 responses





1.3 Problem Definition

Medical images might produce noisy captures or unclear, which could affect the reliability of the results or hide any musculoskeletal conditions, which will be solved by preprocessing. There is no known work which tackles the problems found in the elbow, so we will learn to understand the difficulties faced when diagnosing disorders found in the elbow, in order to be able to extract features (which is not included in previous work) thus enhancing the accuracy of abnormality detection. We will be contributing to further enhancing MURA's dataset by labelling the disorders present in the radiograph. The system we are developing would be referred to as an expert system, since it will have the ability accept new samples and relearn.

2 Project Description

2.1 Objective

There is a noticeable lack of work on detecting MSD's in general and also in specific regions alone. The most recent work does not comply to our needs, too. That is why this system aims to focus on not only accurately detecting the abnormalities in the elbow, but also classifying the present disorder. By allowing the system to accept new samples and relearn from experienced radiologists, we will have an expert system that would also assist other radiologists with their diagnosis. This project will help with the lack of automatic recognition systems on detecting musculoskeletal disorders specifically on elbows, assist radiologists and also provide a dataset that includes images labelled with the disorders, since there are no available datasets with such characteristic.

2.2 Scope

The system will:

1. Extract features from images to help classify disorders.
2. Include genetic algorithms.
3. Improve the accuracy of detecting abnormalities in the elbow.
4. Develop into an expert system by accepting new samples and relearning.

2.3 Project Overview

3 Similar System Information

1. **MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs**

   The model′s pipeline design in this study [1] acquire one or more views for a disorder. For every view, a 169 layered convolutional neural network (CNN) anticipates the probability of anomalies which could be detected, using the arithmetic mean of the anomalies probabilities from CNN. A Dense Convolutional Network architecture was utilized by the network which is essential in connecting each layers to anticipate results beforehand. The final fully connected layer located in the network was substituted by a layer that has a single output, followed by applying a sigmoid nonlinearity. Cohen's kappa statistic was used to assess the model's accuracy compared with the radiologists 's [1]. On some areas as the wrist and the finger, the model achieved a score nearly close to the radiologists'accuracy. While on other parts as the forearm, humerus and shoulder, the model had the lowest score. Nonetheless, there were areas where the model did not achieve as good as the best radiologist has but still have accomplished a better result than the worst radiologist, such as the elbow and hand areas. They used Receiver Operating Characteristic (ROC) to graph the model's specificity against sensitivity in comparison with the radiologists'. After the model predicted the abnormality possibility, a curve had been initialized from changing the threshold values used for the classification boundary. The performance of the radiologists lied above the curve, pointing out that the model

was not as efficient to detect abnormalities relative to radiologists. The accuracy of the model of the elbow was 0.710 which is equivelanet to the worst radiologistâĂŹs performance.

2. **A Pattern Recognition Approach To The Development of A Classification System for Upper-limb Musculoskeletal Disorders of Workers**

   A clustering system was developed to categorize profiles. Participants were asked to label their profiles to build a recognizable pattern of symptoms. The profiles were clustered according to similar symptoms. The clusters were then combined into a consolidated classification system. A clinical expert group helped in testing and evaluating the reliability and the accuracy of the classification system based on approach proposed in [2] and [3]. The responses was assessed using an unweighted kappa statistic [4]. The revision of the consolidated classification system was executed by six clinical experts, and set a goal of 0.75 for the interobserver reliability. The consolidated classification system [5] summarized the findings of the two formal workshops by describing four features; the zone (the number and location of the major findings), type of disorder (pattern of signs and symptoms) and presence of neurological findings and the duration of symptoms. The interobserver reliability gave kappa statistics ranging 0.61 (signs) - 0.73 (symptoms), with an overall kappa of 0.65.

3. **Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-based Approach**

   A Deep Siamese architecture [6] was proposed depending on convolutional neural network (CNN). It detects knee osteoarthritis (OA). Kellgren Lawrence (KL) method [7] was applied to decide severity level. They used a system [**?**] posted before to locate the knee area. In the design of CNN, a small convolutional stack filters of 3x3 is used. Three models are trained with three different seeds. Predictions are summed and transferred to Softmax layer for normalization. Probability of KL j grade

of image x out of k classes is defined by equation 1.

$$P(y=j|\mathbf{x}) = \frac{\exp\left[\sum_{m=1}^{M} \hat{P}_m(y=j|\mathbf{x})\right]}{\exp\left[\sum_{k=1}^{K}\sum_{m=1}^{M} \hat{P}_m(y=k|\mathbf{x})\right]} \quad (1)$$

Training the model on multicenter osteoarthritis study (MOST) dataset containing 3,026 subjects and tested by 3000 random samples of knee OA from Osteoarthritis Initiative (OAI) dataset containing 4,796 subjects. They have reached an accuracy of Quadratic kappa coefficient of 0.83 and average multiclass accuracy of 66.71% .The diagnosis of radiological OA reached AUC (area under the curve) of 0.93 in a ROC (Receiver operator curve) curve.

4. **Segmentation of Pathological Spines in CT Images Using a Two-Way CNN and a Collision-Based Model**

The system [8] introduces a two-way CNN architecture. It was trained on images of normal spines. Image patches were extracted after choosing a center voxel randomly. The patches of size 27x27x27 and 55x55x55 were sent through to two different paths. The first path had nine convolutional layers and residual connections followed by a parametric rectified linear unit (PReLU).The second path had three convolutional layers followed by (PReLU) and a max-pooling layer repeatedly. Finally, a layer of a transposed convolution layer followed by a PReLU is applied to ensure that both paths have the same output size. The outputs were jointly sent to three fully connected layers and finally to a softmax function. Classification is done into three segmentation classes; anterior arch, posterior arch and background, probabilities. The second step was initializing the collision-based model, which was a mesh union of two successive vertebrae (denoted as $M1$ and $M2$) with their computed mean pose and shape. The performance of segmenting vertebrae [8] was measured by computing three different values, the first is Dice similarity coefficient (DSC), and the second is mean symmetric surface distance (MSD)while the third is Hausdorff surface distance (HD). The mean ± standard deviation of the three computed measurements were 93.2±2.2%, 0.5±0.2mm, 8.4±3.4mm, respectively.

5. Shape-Aware Deep Convolutional Neural Network for Vertebrae Segmentation

The UNet architecture [9] is used for extracting different objects. This is achieved in CNN by contracting path to reduce input size from 64x64 to 8x8 and another expanding path. 9 convolutional layers were applied per contracting path. Batch normalization and rectified linear unit layer are then applied. In addition to the 9 layers, another three 2x2 pooling layers were applied. One pooling layer is preceded by two convolutional layers. The proposed approach achieved upsampling then concatenation between data in the expanding path and its corresponding data in the contracting path. A single channel vertebra patch of size 64x64 is extracted by the network, and a two probabilistic output model is used for the prediction of the same vertebra mask. The network has 24,238,210 parameters. A dataset having (x)- segmentation label (y) pairs, with a set of W parameters. This set of parameterms aim to minimize a loss function Lt. The pixel-wise log loss, which is the loss function form and is updated and extended, needed for segmentation is shown in equation 2

$$\hat{W} = arg\min_{\mathbf{W}} \sum_{n=1}^{N} (L_t(\{x^{(n)}, y^{(n)}\}; W) + L_s(\{x^{(n)}, y^{(n)}\}; W)) \tag{2}$$

Where N represent number of training samples, while $x(n), y(n)$ is the n-th training sample with its corresponding manual segmentation. The pixel-wise segmentation loss per sample is computed by equation 3.

$$L_t = (\{x, y\}; W) = -\sum_{i \epsilon \Omega_p} \sum_{j=1}^{M} y_i^j \log P(y_i^j = 1 | x_i; W), \tag{3}$$

$$P(y_i^j = 1 | x_i; W) = \frac{\exp(a_j(x_i))}{\sum_{k=1}^{M} exp(a_k(x_i))}$$

A novel shape-based term, Ls, is introduced for the training of the segmentation network. The network produces prediction

masks from the training vertebral shapes and uses them to predict objects as shown in equation 4.

$$L_s(\{x,y\};W) = -\sum_{i\epsilon\hat{\Omega}_p}\sum_{j=1}^{M} y_i^j E_i \log P(y_i^j = 1|x_i;W); E_i = D(\hat{C}, C_{GT}) \tag{4}$$

UNet-S is the updated segmentation network using the loss function defined by equation 4. Training was applied for 30 epochs and batch of 25 vertebra patches. Testing was applied on 792 vertebrae extracted from 172 test images manually. Evaluation was achieved by a metric called fit failure [10].

6. **Classification of Osteoporotic Vertebral Fractures Using Shape and Appearance Modelling**

Diagnosing vertebral fracture [11] and determining its severity level depends on measuring of vertebral body heights. An approach consist of modeling the format of vertebrates height in addition of k-nearest neighbour and a strong classifier which is random decision forests (RF) for high quality segmentation of partial or full vertebrates of the body. This approach used on plain radiograph and computed tomography (CT). Fracture Size and shape was computed using statistical shape model (SSM) to achieve a linear model. The accuracy improved as the false positive decreased by 60% having sensitivity of 80% . DXA manual annotations the false positive rate (FPR) decreased from 4.4% to 1.7% and the automated decreased from 12.7% to 8.8%. While CT manual annotations decreased from 8.9% to 3.3% , While the automated decreased from 9.7% to 7.0% . Reducing the automatic FPR by $\approx 30\%$ and the manual annotations by $\approx 60\%$.

7. **Reconstruction of 3D Muscle Fiber Structure Using High Resolution Cryosectioned Volume**

The initial phase [12] is the correction of artifact which is a discontinuity of the color balance, discovered throughout the data and clearer in a projection view. Therefore, artifact cor-

rection process was suggested to eliminate this discontinuity. An original image as $h(i, j, k)$ and its transform $H(x, y, z)$ were given and the high frequency area in the z-axis of the transformed image was zero-masked as defined in equation 7.

$$M(x, y, z) = \begin{cases} 0, & \text{if } \sqrt{x^2 + y^2} < t_r \text{ and } |z| > t_z \\ 1, & \text{otherwise} \end{cases}$$

Where M is a zero-valued cylinder along z-axis, tr is radius of the cylinder. The cylinder is filled by ones in the range $-tz < z < tz$. tr and tz were experimented with values 0.2 and 0.08 respectively . The transform of the masked volume $F - 1(H(x, y, z)M(x, y, z))$, shows the effect of artifact correction on volume. Gradient-based structure tensor computes the muscle fiber local orientation around each voxel. Representing Gaussian multiplication respecting both the x and y axis for example. To make the photograph much smoother, another Gaussian multiplication was performed on each element using a standard deviation $\sigma 2$. $\sigma 1$ was determined based on the noise that was contained in the given photograph. After computing the tractography [12] on several muscles in the hips and head regions, the results were the successful identification of each muscle fiber in the two given datasets. The female pelvis dataset showed 24 different muscles around the hips and identifying each muscle with its name, and as for the male whole body dataset it showed 4 masticatory muscles.

8. **Unfolded Cylindrical Projection for Rib Fracture Diagnosis**

The proposed approach [13] introduced a new method for automatic rendering of unfolded clear view of the entire ribcage. Segmentation detected anatomical landmarks using Random Forest (RF) classifier [14] that was trained by histogram of oriented gradients (HOG) [15] features. Using those detected landmarks, they transformed synthetic atlas to novel patient space to bound the segmentation of the ribs by predefining a thorax box where everything outside it is discarded. Finally an algorithm based on demon algorithm [16] was applied. With the resulted field, they warped the three wanted masks: spine,

sternum and ribs and removed other objects, which output a mask with mainly ribs and unrecognized objects like vessels and cables. Removing those vessels done by using a vesselness filter [17]. [13] used a dataset of 369 CT for landmark classifier training, another CT dataset as segmentation atlas for computational efficiency, and 70 for testing. A clinical expert rated the specialized rib view on a Likert scale [13]. The scale values are: Gold - Diagnostic - Moderate - Low Level - Very Low Level of confidence. The testing chest scans were split in 44 known datasets and 26 unknown datasets. The qualitative evaluation resulted in 0.0% gold standard, 62.9% diagnostic confidence, 21.4% moderate confidence, 10.0% low-level of confidence, and 5.7% very low-level of confidence. Six cases from the seen database were manually labelled for a quantitative reference. They were separated as ribs and spine. The manually and automatically generated masks were compared resulting in high DICE, Sensitivity, Specificity values each.

9. **Automatic Localization of the Lumbar Vertebral Landmarks in CT Images with Context Features**

The proposed framework [18] has two modules. One localizes vertebral bodies and estimates their pose by detecting vertebral end plates in 4 steps, while the other localizes key landmarks of vertebrae based on centers and pose predetermined in the first module. Two levels of Random Forest regression technique [19] were applied. During testing, a normally distributed probability map is constructed for every vertebral level and retrieved with the "Mean-Shift Mode-Seeking" algorithm [20]. Second, image deformation was done by curved planar reformation [21]. Re-slicing of CT-scans was done based on the curve that goes through the detected VB centers [22]. The second module was separate localization of vertebral landmarks on each level of the lumbar spine from five levels L5-L1. It had two successive layers of Random Forest regression techniques. The first layer used appearance features while the second layer was applied on context features. Finally, regression was achieved by random forest per landmark [23]. The methods proposed were evaluated on a dataset of 28 CT images. The space between

slices was 0.7 mm, and each slice size was 512x512. All images capture at least the S1-L1 levels, and, in some cases, the thoracic region captures up to level T10. There are five manual annotations made for every lumbar vertebra. One of them were the center of the VB. The other four were computed from the inferior articular process on a lumbar vertebra and include the bottom-left A and the bottom-right B, top-right C and the top-left$D$. In [18], all images are resampled to an isotropic spacing of 1x1x1mm. 20 images are randomly selected for the training phase of the proposed methods. When a VB center detection lies within 10mm from the respective manual annotation, it is considered successful. The mean localization error is 3.2 mm, with a standard deviation of 2.0mm and a median value of 2.8 mm. Overall, the proposed method achieves a mean localization error of 3.0 mm, with a 1.6mm standard deviation and a median value of 2.7 mm. The detections have a localization error under 6 mm represent 95.4%. For the training of the second layer, experiments were done by removing randomly displaced ROIs of [23] and train instead using ROIs centered around the VB centers, which increased the localization error to $3.4 \pm 1.8mm$. The mean dice coefficient, across all the spinal levels, is 88.8%.

10. **3D Cobb Angle Measurements from Scoliotic Mesh Models with Varying Face-Vertex Density**

Detecting the severity level of the deformity or the damage in spine is usually done by measuring the cobb angle. But instead of measuring it from a 2D radiograph, the approach[24] based on measuring it from 3D constructed mesh. 3D model for the spinal cord with alternating density per face to detect spinal deformity scoliosis which is abnormal curve in the spine. Unsupervised machine learning of vertebral column faces is used to locate the center of the vertebral column and detect the faces of the 3D mesh which is used for training.

Then labeling and classifying contiguous faces of the mesh through Principle component analysis (PCA) to convert the vertices from image to the mesh. Identifying and calculating the mesh faces. Considering that the vertebral shape is cylin-

drical, analyzing it to 3 groups. The superior endplate $F_{B,TOP}$, the inferior endplate $F_{B,BOTTOM}$ and the walls representing the vertebral surface $F_{B,SIDE}$. Unsupervised machine learning using k-means++ to classify the highest value cluster to $F_{B,TOP}$ and the lowest value cluster to $F_{B,BOTTOM}$. Then the faces of the mesh are classified into $F_{VB,TOP}$, $F_{VB,BOTTOM}$ and $F_{VB,SIDE}$. Linear classifier is responsible for grouping the contiguous faces of the same class and the group with the maximum number of faces is chosen to be the superior and interior endplates.

$$\begin{aligned} F_{SUP} &= \arg\max_{F_{A,j}} \left(\text{count}\left(F_{A,j} \cap F_{B,TOP}\right)\right), \quad j = 1, 2, \ldots, J \\ F_{INF} &= \arg\max \left(\text{count}\left(F_{A,j} \cap F_{B,BOTTOM}\right)\right), \quad j = 1, 2, \ldots, J \end{aligned} \tag{5}$$

The 3D cobb angle is generated by constructing the superior endplate to the upper part of vertebrae and the inferior to the lower part of the vertebrae in which the deformity located. The parameters of these planes are calculated using the algorithm of random sample consensus (RANSAC) on both faces superior and inferior, ending up with normal vectors. Testing is done on 60 mesh models having adolescent idiopathic scoliosis (spinal deformity) triangular shaped with 48.8° mean cobb angle, in which each model is constructed from 17 varying face densities. Measuring the cobb angle with this semi-automated method even that the upper-end and lower-end of vertebral have to be manually selected, reaching an accuracy on faces having mean edge less than 6mm of 3.0° absolute mean error which is equivalent to 2.2° standard deviation.

11. **Automatic Full Femur Segmentation from Computed Tomography Data-sets Using an Atlas-Based Approach**

The target [25] is to demonstrate the automatic segmentation of the femur in clinical CTs to fully evaluate the health of the bone and to identify the impact of osteoporosis. First, pre-processing by re-sampling the CT scans to a resolution of 0.625mm by utilizing cubic interpolation is achieved. Afterwards, the scans volume is divided into two different sets, one include the right femur while the other include the left femur. Semi-automatic

segmentation is done by graph-cut approach. In this approach, foreground (femur) and background (muscle, pelvis and tibia) regions were identified. By using MITK-GEM, this resulted in the creation of joint space in the segmented images. Therefore, the 6 MITK-GEM segmentations were afterwards manually improved and corrected by the aid of medical professionals. The MITK-GEM [25] approach was the best one so far that obtained a similarity coefficient of 0.99 and a Hausdorff distance of 4.5mm. In atlas-based approach, after eliminating the contralateral femur which resulted in the improvement of both the average Hausdorff distance and dice similarity coefficient. The average Hausdorff distance increased from 5.95mm to 8.53mm and the dice similarity coefficient decreased from 0.978 to 0.969 which is considered better results that they were.

3.1 Similar System Description

The most recent work proposed by P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R. L. Ball is a large dataset containing X-rays of bones and has a system that determines whether an X-ray study is normal or abnormal. Their baseline uses a 169-layer convolutional neural network to detect and localize abnormalities. The baseline is evaluated on the Cohen's kappa statistic and is compared to the radiologist's performance. The model performance scored best on wrist and finger studies.

3.2 Comparison with Proposed Project

Table 1: Comparison

Specification	Dataset	Results
Osteoarthritis in knees [26]	MR Imaging from the Osteoarthritis Initiative where 90 had one or more cartilage signal abnormality at baseline and 90 who did not have any signal abnormalities at baseline for comparison.	After 48 months, 57.1% with signal abnormalities at baseline developed a morphological defect while only 4% of compartments without baseline signal abnormalities developed morphological defects.
Upper-limb disorders [5]	242 workers from a newspaper filled out a questionnaire that has a pain diagram.	Interobserver reliability 0.7
Spinal disorders [8]	10 axially reconstructed CT images of the lumbar spine and thoracolumbar spine each for training and 15 axially reconstructed CT images of the lumbar spine for testing acquired from SpineWeb online platform.	DSC of 93.2±2.2%, MSD of 0.5±0.2mm and HD of 8.4±3.4mm
Lateral cervical spine [9]	296 real-life emergency room lateral cervical X-ray images. Trained on an augmented dataset of 26370 vertebrae, and tested on 792	The proposed framework had an average error of 1.11 pixels, and the introduction of the novel shape-aware term in the loss function scored an average error of 0.99 pixel
Rib fractures [13]	Clinically acquired, 369 CT datasets for landmark classifier training and 70 CT studies for testing.	The qualitative evaluation resulted in 0.0% gold standard, 62.9% diagnostic confidence, 21.4% moderate confidence, 10.0% low-level of confidence, and 5.7% very low-level of confidence.
Knee Osteoarthritis Diagnosis [27]	Training the model on multicenter osteoarthritis study (MOST) dataset containing 3,026 subjects and tested by 3000 random samples of knee OA from Osteoarthritis Initiative (OAI) dataset containing 4,796 subjects.	Quadratic kappa coefficient of 0.83 and average multiclass accuracy of 66.71% .The diagnosis of radiological OA reached AUC (area under the curve) of 0.93 in a ROC (Receiver operator curve) curve.

Osteoporotic Vertebral Fractures [11]	PACS collected CT scans for 868 patients of age 18 and older including thoracic (mid back) and lumbar (lower back) vertebrae scans. Also another list of CT scans collected by PACS for 132 patients of age 60 and older including osteoporotic vertebral fractions. In Digital Imaging and Communications in Medicine (DICOM) format. Trained on 402 volumes containing 132 image of fractures and the rest of images used for validation.	The accuracy improved as the false positive decreased by 60% having sensitivity of 80%
Adolescent idiopathic scoliosis [24]	Testing is done on 60 mesh models having adolescent idiopathic scoliosis of spine triangular shaped with 48.8 mean cobb angle, each model constructed from 17 varying face densities.	On meshes having faces of mean edge less than 6mm they reached an accuracy of 3.0° absolute mean error which is equivalent to 2.2° standard deviation.
Lumbar vertebral bodies [18]	28 lumbar-focused computed tomography images	Detection of landmarks with a mean localization error of 3.0 mm
Body composition [28]	20 CT scans for training and 40 CT scans for testing, obtained from two clinical datasets.	Compared with the manually labelled slices, their method achieved very high spearman's rank correlation coefficients, and higher dice similarity index when compared with the baseline method. [29]
Intervertebral discs [**?**]	MICCAI 2015 IVD segmentation challenge datasets	Mean Dice overlap coefficient: 92.0%, mean average symmetric surface distance: 0.41 mm
Vertebral fractures [30]	Three datasets were used; dataset 1 is composed 45 patients between the ages 25 and 69 and lacking any fractures whatsoever, dataset 2 is composed of 11 patients with vertebral fractures, and finally dataset 3 includes CSI2014 dataset that contains 20 computed tomography (CT).	The Attention-Net aided the Segmentation-Net by filtering stray segmentations, enhancing the dice coefficients (DICE) by 10%.

Femoral fractures [25]	Three cadavers were scanned and resulted in the important bilateral hip CT scans. Resolution was 0.684x0.684mm with a thickness of 0.625mm. The CT scans contain the entire femur from the top of the iliac crest of the pelvis and ends in the proximal tibia.	The MITK-GEM approach was the best one so far that obtained a similarity coefficient of 0.99 and a Hausdorff distance of 4.5mm.
Hip joint and head regions [12]	The Visible Korean Human which includes several images of cryosection of human corpses.	The successful identification of each muscle fiber in the two given datasets. The female pelvis dataset showed 24 different muscles around the hips and identifying each muscle with its name, and as for the male whole body dataset it showed 4 masticatory muscles.
Torso skeletal abnormalities [31]	200 CT torso examinations that were performed in a level one trauma center.	82% of the total number of examinations were established to have at least one finding. A total number of 433 findings was discovered from these examinations and classified into 20% high, 58% moderate or 58% low severity skeletal findings.
Upper-limb extremity [1]	MURA has 9,045 normal and 5,818 abnormal musculoskeletal radiographic studies of the upper extremity.	On some areas, the model's performance was comparable to the radiologists', while in other areas it still needed further improvement.

4 Project Management and Deliverables

4.1 Tasks and Time Plan

Task	Delivery
Dataset	1 working month
Application	4 working months
Cloud launch	1 month
SRS	October
SDD	February

4.2 Budget and Resource Costs

The only item that would be purchased is a PC.

References

[1] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R. L. Ball, et al., Mura dataset: Towards radiologist-level abnormality detection in musculoskeletal radiographs.

[2] H. C. Kraemer, A. F. Korner, Statistical alternatives in assessing reliability, consistency, and individual differences for quantitative measures: Application to behavioral measures of neonates., Psychological Bulletin 83 (5) (1976) 914–921.

[3] A. Donner, M. Eliasziw, Sample size requirements for reliability studies, Statistics in medicine 6 (4) (1987) 441–448.

[4] A. J. Viera, J. M. Garrett, et al., Understanding interobserver agreement: the kappa statistic, Fam Med 37 (5) (2005) 360–363.

[5] D. E. Beaton, C. Bombardier, D. C. Cole, S. Hogg-Johnson, D. Van Eerd, C. E. Group, A pattern recognition approach to the development of a classification system for upper-limb musculoskeletal disorders of workers, Scandinavian journal of work, environment & health (2007) 131–139.

[6] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, S. Saarakkala, Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach, Scientific reports 8 (1) (2018) 1727.

[7] A. Tiulpin, J. Thevenot, E. Rahtu, S. Saarakkala, A novel method for automatic localization of joint area on knee plain radiographs, in: Scandinavian Conference on Image Analysis, Springer, 2017, pp. 290–301.

[8] R. Korez, B. Likar, F. Pernuš, T. Vrtovec, Segmentation of pathological spines in ct images using a two-way cnn and a collision-based model, in: International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer, 2017, pp. 95–107.

[9] S. M. R. Al Arif, K. Knapp, G. Slabaugh, Shape-aware deep convolutional neural network for vertebrae segmentation, in: International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer, 2017, pp. 12–24.

[10] P. Bromiley, J. Adams, T. Cootes, Localisation of vertebrae on dxa images using constrained local models with random forest regression voting, in: Recent Advances in Computational Methods and Clinical Applications for Spine Imaging, Springer, 2015, pp. 159–171.

[11] P. A. Bromiley, E. P. Kariki, J. E. Adams, T. F. Cootes, Classification of osteoporotic vertebral fractures using shape and appearance modelling, in: International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer, 2017, pp. 133–147.

[12] Y. Otake, K. Miyamoto, A. Ollivier, F. Yokota, N. Fukuda, L. J. OâĂŹDonnell, C.-F. Westin, M. Takao, N. Sugano, B. S. Chung, et al., Reconstruction of 3d muscle fiber structure using high resolution cryosectioned volume, in: International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer, 2017, pp. 85–94.

[13] C. Tobon-Gomez, T. Stroud, J. Cameron, D. Elcock, A. Murray, D. Wyeth, C. Conway, S. Reynolds, P. A. G. Teixeira, A. Blum, et al., Unfolded cylindrical projection for rib fracture diagnosis, in: International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer, 2017, pp. 36–47.

[14] A. Criminisi, J. Shotton, Decision forests for computer vision and medical image analysis, Springer Science & Business Media, 2013.

[15] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 1, IEEE, 2005, pp. 886–893.

[16] W. Crum, L. Griffin, D. Hill, D. Hawkes, Zen and the art of medical image registration: correspondence, homology, and quality, NeuroImage 20 (3) (2003) 1425–1437.

[17] A. F. Frangi, W. J. Niessen, K. L. Vincken, M. A. Viergever, Multiscale vessel enhancement filtering, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 1998, pp. 130–137.

[18] D. Damopoulos, B. Glocker, G. Zheng, Automatic localization of the lumbar vertebral landmarks in ct images with context features, in: International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer, 2017, pp. 59–71.

[19] B. Glocker, D. Zikic, E. Konukoglu, D. R. Haynor, A. Criminisi, Vertebrae localization in pathological spine ct via dense classification from sparse annotations, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, pp. 262–270.

[20] Y. Cheng, Mean shift, mode seeking, and clustering, IEEE transactions on pattern analysis and machine intelligence 17 (8) (1995) 790–799.

[21] A. Kanitsar, D. Fleischmann, R. Wegenkittl, P. Felkel, M. E. Gröller, Cpr: curved planar reformation, in: Proceedings of the conference on Visualization'02, IEEE Computer Society, 2002, pp. 37–44.

[22] A. Kanitsar, D. Fleischmann, R. Wegenkittl, P. Felkel, M. E. Gröller, Cpr: curved planar reformation, in: Proceedings of the conference on Visualization'02, IEEE Computer Society, 2002, pp. 37–44.

[23] Y. Gao, D. Shen, Context-aware anatomical landmark detection: application to deformable model initialization in prostate ct images, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2014, pp. 165–173.

[24] K. R. P. S. K. S. . V. T. PetkoviÄĞ, U., 3d cobb angle measurements from scoliotic mesh models with varying face-vertex density, in: International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer, 2017, pp. 48–58.

[25] B. A. Besler, A. S. Michalski, N. D. Forkert, S. K. Boyd, Automatic full femur segmentation from computed tomography datasets using an atlas-based approach, in: International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer, 2017, pp. 120–132.

[26] B. J. Schwaiger, A. S. Gersing, J. Mbapte Wamba, M. C. Nevitt, C. E. McCulloch, T. M. Link, Can signal abnormalities detected with mr imaging in knee articular cartilage be used to predict development of morphologic cartilage defects? 48-month data from the osteoarthritis initiative, Radiology 281 (1) (2016) 158–167.

[27] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, S. Saarakkala, Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach, Scientific reports 8 (1) (2018) 1727.

[28] P. Hu, Y. Huo, D. Kong, J. J. Carr, R. G. Abramson, K. G. Hartley, B. A. Landman, Automated characterization of body composition and frailty with clinically acquired ct, in: International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer, 2017, pp. 25–35.

[29] W. Zhang, J. Liu, J. Yao, R. M. Summers, Segmenting the thoracic, abdominal and pelvic musculature on ct scans combining atlas-based model and active contour model, in: Medical Imaging 2013: Computer-Aided Diagnosis, Vol. 8670, International Society for Optics and Photonics, 2013, p. 867008.

[30] A. Sekuboyina, J. Kukačka, J. S. Kirschke, B. H. Menze, A. Valentinitsch, Attention-driven deep learning for pathological spine segmentation, in: International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer, 2017, pp. 108–119.

[31] J. W. Kung, J. S. Wu, S. K. Shetty, V. C. Khasgiwala, P. Appleton, M. G. Hochman, Spectrum and detection of musculoskeletal findings on trauma-related ct torso examinations, Emergency radiology 21 (4) (2014) 359–365.