[https://www.nature.com/articles/s41598-021-88578-w](https://www.nature.com/articles/s41598-021-88578-w" \t "_blank)  
<https://arxiv.org/abs/2010.12030>  
<https://www.researchgate.net/publication/333788108_Abnormality_Detection_in_Musculoskeletal_Radiographs_Using_Capsule_Network>  
<https://ieeexplore.ieee.org/document/8736807>  
<http://www.icrjournal.ir/article_100902.html>  
<https://www.semanticscholar.org/paper/Musculoskeletal-radiographs-classification-using-Harini-Ramji/aa7ec5dde0a5286d7e8d112130b7744ea2f4e734>  
<https://amrita.edu/publication/musculoskeletal-radiographs-classification-using-deep-learning/>  
<https://www.auntminnie.com/index.aspx?sec=ser&sub=def&pag=dis&ItemID=133396>  
<https://github.com/Youssefares/MURA-Abnormality-Detection-in-Musculoskeletal-Radiographs>  
<https://www.readcube.com/articles/10.3390%2Freports2040026>

IEEE:

**Deep CNN-Based Ensemble CADx Model for Musculoskeletal Abnormality Detection from Radiographs**

<https://ieeexplore.ieee.org/document/8975455>

# Abnormality Detection in Musculoskeletal Radiographs using EfficientNets

<https://ieeexplore.ieee.org/document/9375275>

# KALM: Key Area Localization Mechanism for Abnormality Detection in Musculoskeletal Radiographs

<https://ieeexplore.ieee.org/document/9053768>

# Abnormality Detection in Humerus Bone Radiographs Using DenseNet

<https://ieeexplore.ieee.org/document/9696904>

# Abnormality Detection in Musculoskeletal Radiographs Using Capsule Network

<https://ieeexplore.ieee.org/document/8736807>

# Deep Learning for Musculoskeletal Image Analysis

<https://ieeexplore.ieee.org/document/9048671>

# Branding - Fusion of Meta Data and Musculoskeletal Radiographs for Multi-Modal Diagnostic Recognition

<https://ieeexplore.ieee.org/document/9022072>

<https://www.nature.com/articles/s41598-021-88578-w>

<https://arxiv.org/pdf/2010.12030.pdf>

<https://arxiv.org/ftp/arxiv/papers/1908/1908.02170.pdf>

# MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs

<https://arxiv.org/abs/1712.06957>

# Bone Fracture Detection and Localization on MURA Database Using Faster-RCNN

<https://ieeexplore.ieee.org/document/9729393>

# Self-Taught Semi-Supervised Anomaly Detection On Upper Limb X-Rays

<https://ieeexplore.ieee.org/document/9433771>

# Radiography Classification: A Comparison between Eleven Convolutional Neural Networks

<https://ieeexplore.ieee.org/document/9264285>

# Radiography Classification: A Comparison between Eleven Convolutional Neural Networks

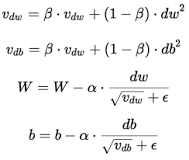
 Eleven convolutional neural network architectures (GoogleNet, Vgg-19, AlexNet, SqueezeNet, ResNet-18, Inception-v3, ResNet-50, Vgg-16, ResNet-101, DenseNet-201 and Inception-ResNet-v2) were used to classify a series of x-ray images from Stanford Musculoskeletal Radiographs (MURA) dataset corresponding to the wrist images of the data base. For each architecture, the results were compared against the known labels (normal / abnormal) and then the following metrics were calculated: accuracy (labels correctly classified) and Cohen's kappa (a measure of agreement) following MURA guidelines. Numerous experiments were conducted by changing classifiers (Adam, Sgdm, RmsProp), the number of epochs, with/without data augmentation. The best results were provided by InceptionResnet-v2 (Mean accuracy = 0.723, Mean Kappa = 0.506)

Pre processing

1. **Pre-processing steps,** which may consist of: Low pass filtering to remove high-frequency noise, cropping of images to remove excessive background region (notice that some of the incorrect classifications in Fig. 3 had large background regions). More elaborate preprocessing approaches such as location and orientation of bones [35] could help detect the areas of real interest, and discard any region that may be biasing results, such as the labels for right or left hand, which being always very bright might be confusing the architectures.

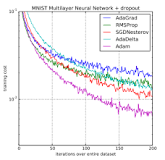
* Figure 4A-4C above shows spatial and frequency domain images and their obtained histogram of Catphan700 phantom. Figure 4A show the original image and histogram with a calculated noise level of standard deviation as high as 49.937. The application of a low pass median filter in the spatial medium of radius 2.0 pixels reduces the noise level slightly by 0.824% to 49.527 as shown in Figure 4B. However, when the low pass filter is applied to the original image in the frequency domain of the same radius (Figure 4C), the noise level is reduced sharply by 88.74% from 49.937 to 19.243. There are several different types of filters that can be applied in noise reduction in digital radiography. These include median filter, FIR filter, Gaussian filter, mode filter, convolver, mean, unsharp mask and others. The type of filter to apply in noise reduction is based on the physical appearance upon the application [7]. Also, the choice of filter for reducing noise in medical images depends on the type of noise and type of filtering technique. Among all these linear filters, the Gaussian filter is the most common as it plays an important role in both theory and applications. However, Gaussian denoising filter is known to over smooth images resulting in loss of significant detail, most importantly edge sharpening [12].
* Noise in a tomography image can be reduce through reconstruction. The reconstruction process uses two main methods. These include analytical method (Filtered Back Projection-FBP) and iterative reconstruction method [17].
* The common types of noise that are present in x-ray images (Figure 3A & 3B) are Poisson noise, salt and pepper noise, and speckle noise.
* In conventional x-ray techniques, there are two main sources of noise. The primary source of these noise is the secondary radiation which is produced as a result of scattered radiation from the x-ray machine and scattered radiation by an object which reach the film. When the useful x-ray beam is intercepted by any object, a secondary scattered radiation is produced.
* The manner in which the film is processed and handled is another way of introducing noise into the final image. This leads to film retake and over exposure of patients to unnecessary radiation. The use of specially designed grid plates made of lead have proven to reduce the scattered secondary x-ray radiation and produce radiographs within acceptable noise level [7]. In a situation where the film is under processed, the sensitivity and contrast is reduced below specific values.
* poison noise is as a result of uneven distribution of x-rays over the receptor surface
* Speckle noise on the other hand occurs as a granular appearance in an image which is produced as a result of random fluctuations in the return signal from an object which is not found to be bigger than a single image processing element. Within a specific area, speckle noise is able to increase the mean grey level
* Low pass (smoothing) filters are applied to the images to remove the noise in them. These filters which are applied in either spatial or frequency domain have the ability to reduce contrast among neighboring pixels and thereby reducing the amount of noise in the image [7]

What is RMSProp Optimizer in deep learning?

[[](https://www.google.com/search?rlz=1C1CHBF_enIN891IN891&sxsrf=APq-WBup-LX3KAPOzxW3ZoDUc0qLoz0oZQ:1648610353979&q=What+is+RMSProp+Optimizer+in+deep+learning?&tbm=isch&source=iu&ictx=1&vet=1&fir=rg-5DH3uMgZZ6M%252CtanArST7QIUQeM%252C_&usg=AI4_-kTg8nc5Ox85Vrr1Vs7kBcdFG-soxQ&sa=X&ved=2ahUKEwjDipaU8Oz2AhWbR2wGHWd5Dt8Q9QF6BAgWEAE#imgrc=rg-5DH3uMgZZ6M)](https://www.google.com/search?rlz=1C1CHBF_enIN891IN891&sxsrf=APq-WBup-LX3KAPOzxW3ZoDUc0qLoz0oZQ:1648610353979&q=What+is+RMSProp+Optimizer+in+deep+learning?&tbm=isch&source=iu&ictx=1&vet=1&fir=rg-5DH3uMgZZ6M%252CtanArST7QIUQeM%252C_&usg=AI4_-kTg8nc5Ox85Vrr1Vs7kBcdFG-soxQ&sa=X&ved=2ahUKEwjDipaU8Oz2AhWbR2wGHWd5Dt8Q9QF6BAgWEAE" \l "imgrc=rg-5DH3uMgZZ6M)

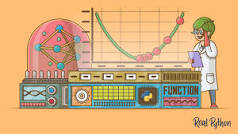
RMSprop is **a gradient-based optimization technique used in training neural networks**. This normalization balances the step size (momentum), decreasing the step for large gradients to avoid exploding and increasing the step for small gradients to avoid vanishing. ...

What is Adam optimization in deep learning?

[[](https://www.google.com/search?rlz=1C1CHBF_enIN891IN891&sxsrf=APq-WBup-LX3KAPOzxW3ZoDUc0qLoz0oZQ:1648610353979&q=What+is+Adam+optimization+in+deep+learning?&tbm=isch&source=iu&ictx=1&vet=1&fir=kZ03liYln8Me7M%252Cj_pGhnPLJnadwM%252C_&usg=AI4_-kSugZX28UkMazE2YS1-SRSa68I_Bg&sa=X&ved=2ahUKEwjDipaU8Oz2AhWbR2wGHWd5Dt8Q9QF6BAgXEAE#imgrc=kZ03liYln8Me7M)](https://www.google.com/search?rlz=1C1CHBF_enIN891IN891&sxsrf=APq-WBup-LX3KAPOzxW3ZoDUc0qLoz0oZQ:1648610353979&q=What+is+Adam+optimization+in+deep+learning?&tbm=isch&source=iu&ictx=1&vet=1&fir=kZ03liYln8Me7M%252Cj_pGhnPLJnadwM%252C_&usg=AI4_-kSugZX28UkMazE2YS1-SRSa68I_Bg&sa=X&ved=2ahUKEwjDipaU8Oz2AhWbR2wGHWd5Dt8Q9QF6BAgXEAE" \l "imgrc=kZ03liYln8Me7M)

Adam is **a replacement optimization algorithm for stochastic gradient descent for training deep learning models**. Adam combines the best properties of the AdaGrad and RMSProp algorithms to provide an optimization algorithm that can handle sparse gradients on noisy problems.03-Jul-2017

What is stochastic gradient descent?

[[](https://www.google.com/search?rlz=1C1CHBF_enIN891IN891&sxsrf=APq-WBuI-LsV3RQ1IbJdYZmb5gK85T6YhQ:1648610815077&q=What+is+stochastic+gradient+descent?&tbm=isch&source=iu&ictx=1&vet=1&fir=ODXy6kOHbYPl2M%252CfUgqwsGqMOWnxM%252C_&usg=AI4_-kTuoo6QI6fLgSe92nWZCoNfXWzOpw&sa=X&ved=2ahUKEwiMo4Xw8ez2AhWbR2wGHWd5Dt8Q9QF6BAgbEAE#imgrc=ODXy6kOHbYPl2M)](https://www.google.com/search?rlz=1C1CHBF_enIN891IN891&sxsrf=APq-WBuI-LsV3RQ1IbJdYZmb5gK85T6YhQ:1648610815077&q=What+is+stochastic+gradient+descent?&tbm=isch&source=iu&ictx=1&vet=1&fir=ODXy6kOHbYPl2M%252CfUgqwsGqMOWnxM%252C_&usg=AI4_-kTuoo6QI6fLgSe92nWZCoNfXWzOpw&sa=X&ved=2ahUKEwiMo4Xw8ez2AhWbR2wGHWd5Dt8Q9QF6BAgbEAE" \l "imgrc=ODXy6kOHbYPl2M)

Stochastic gradient descent is **an optimization algorithm often used in machine learning applications to find the model parameters that correspond to the best fit between predicted and actual outputs**. It's an inexact but powerful technique. Stochastic gradient descent is widely used in machine learning applications.

Batch gradient descent means training whole dataset. Mini batch gradient descent means training small chunks of dataset. Stochastic gradient descent means training 1 sample per epoch

What is Adagrad optimizer?

Adagrad is **an optimizer with parameter-specific learning rates, which are adapted relative to how frequently a parameter gets updated during training**. The more updates a parameter receives, the smaller the updates.

1. **Pre-processing steps,** which may consist of: Low pass filtering to remove high-frequency noise, cropping of images to remove excessive background region (notice that some of the incorrect classifications in Fig. 3 had large background regions). More elaborate preprocessing approaches such as location and orientation of bones [35] could help detect the areas of real interest, and discard any region that may be biasing results, such as the labels for right or left hand, which being always very bright might be confusing the architectures.
2. **Post-processing steps** may also be considered, for instance, the association between key features and the predicted classes [32], [43]. Furthermore, the visualisation of key features may be useful to stakeholders (e.g. clinicians or radiologists) who might be more interested in the attributes of the original data rather than the architectures themselves [31].
3. **Ensembles** or combination of different configurations may also help increase the results of individual configurations.
4. Finally, adding **domain knowledge** in terms of knowledge of the anatomical region (i.e. elbow or hand) with the possible cases (i.e. fracture or implant) may allow the fine tuning of the architectures to detect not only an abnormality but the type of abnormality and the location of this.

**67**