Computer-Aided Diagnosis System for Bone Fracture Detection and Classification: A Review on Deep Learning Techniques

*Chapter*

# COMPUTER AIDED DIAGNOSIS SYSTEM FOR BONE FRACTURE DETECTION AND CLASSIFICATION: A REVIEW ON DEEP LEARNING TECHNIQUES

***Leonardo Tanzi** [*, 1], **Enrico Vezzetti** [1], **Alessandro Aprato** [2], **Andrea Audisio** [2], **Alessandro Massè** [2]*

[1]Department of Management and Production Engineering, Politecnico di Torino, Torino, Italy;

[2]School of Medicine, University of Turin, Torino, Italy

## ABSTRACT

Bone fracture detection and classification was a large discussed topic over the last few years and many researchers proposed different technological solutions to tackle this task. Despite this, a universal approach able to support the classification of fractures in the human body still does not exist today. We aim to provide a first discussion concerning a selection of research works done in the technological domain, with a specific focus on Deep Learning. The objective was to underline a picture on the most promising studies for stimulating a knowledge improvement in the specific focus of bone fracture classification, necessary to start the development of an optimal shared framework. The evaluation has been made involving a first qualitative assessment based on strengths and weaknesses, providing a usage scenario evaluation. This could support the development of a helpful Computer Aided Diagnosis (CAD) system able to drive doctors in diagnosis tasks reducing diagnosis time, especially in the most complex tasks, and supporting the reduction of wrong diagnosis issues, especially during stressful working conditions, as what frequently happens in many emergency departments.

---

* Corresponding Author address
Email: leonardo.tanzi@polito.it

**Keywords**: Deep Learning, Neural Networks, Bone Fracture, CAD System

## 1. INTRODUCTION

In the last decades, several medical procedures has improved year by year with the help of technology [1,2]; among these, great progress has been made in the orthopedic field. Bones' fractures are one of the most common injuries nowadays. Every year 2.7 million fractures occur across the EU6 nations, France, Germany, Italy, Spain, Sweden, and the UK [3], an incredible number of people suffers for this disorder and the implications of an untreated fracture may lead to permanent damages or even death. A great responsibility in this lies with the doctors, who have to evaluate tens of X-Ray images a day. The technology utilized for first diagnosis is mostly X-Ray, which is a modality used for more than one hundred years and is still frequently used. It is challenging for doctors to evaluate X-Ray images: firstly, X-Ray could hide certain particularities of bone; secondly, a long experience is needed to correctly classify different types of fractures; thirdly, doctors have often to act in emergency situations and may be constrained by fatigue. Actually, it has been shown that performance of radiologist in the interpretation of musculoskeletal radiographs decrease in fracture detection at the end of the work day compared to beginning of work day [4]. In addition, radiographic interpretation often takes place in environments without the availability of qualified colleagues for second opinions [5]. The success of the treatment and prognosis strongly depends on an accurate classification of the fracture among standard types, such as those defined by the Arbeitsgemeinschaft für Osteosynthesefragen (AO-foundation). In that context, computed aided diagnosing system seem to be able to help doctors especially in very critical scenario, as for instance in emergency department. Deep Learning resulted as the most performative solution for this task, for this reason we decide to focus on this paradigm, underlying the main aspects behind Deep Learning application in the next section.

## 2. DEEP LEARNING SCENARIO

Unfortunatly at present a univoque and shared methodology in bone fracture classification is still not available, but many different studies are proposed expecially in the domain on machine learning domain. Machine Learning is generally defined as the practice of using algorithms to parse data, learn from it, and then make a prediction about something in the world. Two types of machine learning can be defined (we don't consider reinforcement learning for the sake of simplicity): supervised, where the network learns from labelled data, and unsupervised, where the computer learns by itself without any help from the labelling. Deep Learning is becoming more and more widely used in the world of computer vision technologies, giving astonishing results in different fields of application, for example surgery [6] and face recognition [7]. A neural network, i.e the classic architecture used in Deep Learning application, is composed of input, hidden and output layers, all of which are composed of nodes. All these subsequent layers define a function with thousands or even millions of parameters (called weights and biases). The input layer takes in a numerical representation of data (e.g. images with pixel specs), the output layer output predictions, while the hidden layers are correlated with most of the computation. In a typical classification problem, the network is fed with the input images and tries to assign a specific pre-defined class to each of them. After the prediction, it calculates a loss function (how much the predicted classes differ from the original) and adjust its parameters with gradient descent and backpropagation. The images are fed different times through the network back and forth, until the network gets a certain accuracy. In order to be clear, we're going to define some basic Deep Learning glossary.

*Train, Validation, Test Set* It's a common procedure to divide the dataset in 3 different groups: Train Set, to train the network, Validation Set, to test the network after each epoch and Test Set, to test the network at the end of the training with images that it has never seen before.

*Transfer Learning* It has been demonstrated that the parameters of a network trained on some task using a dataset X, may be re-used to adapt the network to solve a different task using a dataset Y, instead of using randomly initialized parameters.

*Data augmentation* Data augmentation is a strategy that enables to significantly increase the diversity of data available for training models,

Figure 1. Examples of data augmentation techniques.

without actually collecting new data, using techniques such as cropping, padding, and flipping. Some examples are showed in Figure 1.

*Convolutional Neural Network (CNN)* CNN are the most used technology in computer vision as they're able to successfully capture the spatial and temporal dependencies in an image through the application of relevant filters, saving memory space and limiting the number of parameters to compute. The input images are made to convolute with different filters to extract feature and then passed to pooling layer to lower the size of the data. After this process the output is passed to the fully connected layers as it's showed in Figure 2.

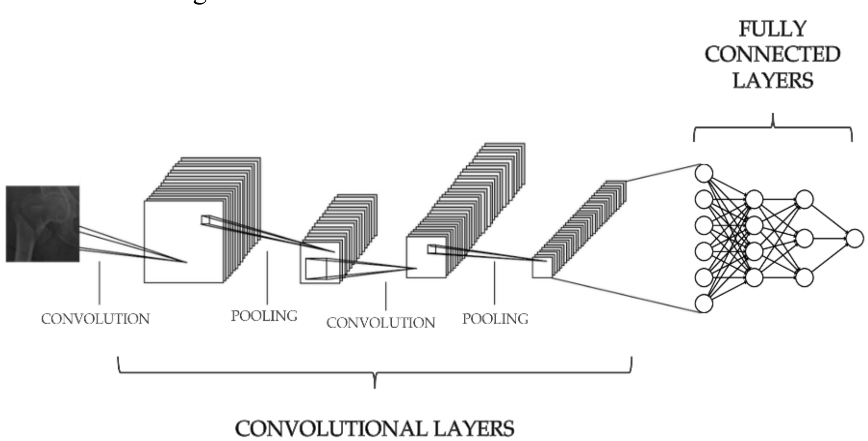


Figure 2. Convolutional Neural Network architecture. Convolutional layers are the layers where filters are applied to the original image, or to other feature maps in a deep CNN. Pooling layers' function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network.

## 3. METHODOLOGY

The review process proposed in the present paper was based on a desktop research analysis [8] run on a selection of specific papers of the specific research domain moving from basic approaches to the main advanced solutions. Initial prior works for detection and classification of fractures [9–11] have focused on conventional machine learning processes consisting of pre-processing, feature extraction and classification steps. Recently, impressive results have been obtained using Deep Learning [12] methods. We summarized a brief overview about Deep Learning in section 1.1. A total of 237 records were identified through database searching and other sources. 107 records were screened and 63 of them excluded, resulting in 44 full-text articles assessed for eligibility. Among these papers, we selected 10 records for analysis. We excluded 21 studies using surpassed technologies and 13 studies tackling non-ineherent arguments. The majority of them pursue the classification between fractured and not fractured bones, while just two of them tried to classify the different types of fractures. We have chosen papers which, in our personal opinion, contain strengths given by a Deep Learning approach that should be used in order to develop a generic tool able to classify every type of fracture in each bone of the human body.

## 4. APPROACHES

To the best of our knowledge, a work that tries to define an ideal method to classify fractures valid for each bone in the human body does not exist. In our opinion, the best way to pursue this is to evaluate different papers and select the strengths which could be mixed together to define a baseline approach. Before the advent of CNN, the pre-processing phase was a fundamental part of the work. For example, in the work of Dimililer [13], the author's aim was to classify whether a bone in an X-Ray image is fractured or not. The system is composed by a neural network following a pre-processing phase. The tool has been trained with 30 images and tested with 70. The images contain different fractures in size and illumination conditions for each subject. In the pre-processing phase, the images are processed using techniques such as Haar Wavelet and Scale-Invariant Feature Transform (SIFT). Haar Wavelet transform is needed to pre-process images in order to compress them and save memory space, SIFT is a powerful method to detect feature points with high resilience to several issues like rotation, compression, and scaling. In the classification phase the author implements a 3-layers neural network with 1024 input neurons. The whole pipeline is showed in Figure 3.

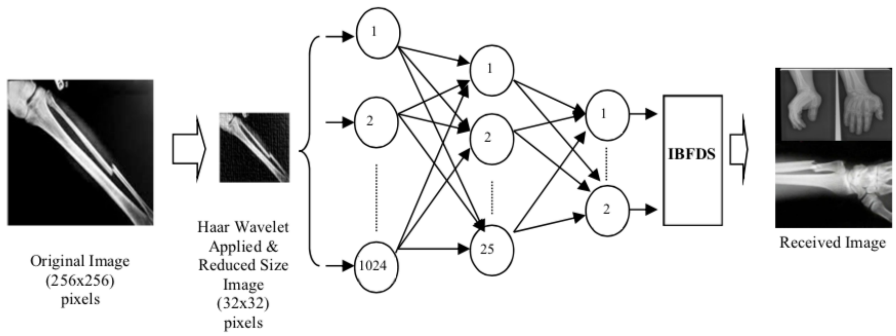


Figure 3. Pipeline of the architecture proposed by Dimililer [13].

This paper describes the technology used before the advent of the convolutional layers. The pre-processing phase was a fundamental part of the process in order to feed the fully connected layers with the correct information. All the remaining papers use convolutional layers to extract features before feeding the fully connected layers.

A dataset of 100 images is quite scant for Deep Learning applications: to solve this problem, Kim and MacKinnon [14] proposed the implementation of transfer learning techniques in order to classify wrist fractures in two classes: broken and unbroken. According to the authors, this was the first work were transfer learning from pre-trained CNNs has been successfully applied to the problem of fracture detection on plain radiographs. The final dataset was composed of 1.389 images, 695 wrist radiographs showing a fracture and 694 showing no fracture. Subsequently, data augmentation was applied: horizontal flip, rotation (between 0° and 25°), width and height shift (by a factor of 0-15%), shearing (between 0-10%) and zoom (between 0-15%). This resulted in an overall amplification by a factor of 8 and 5.560 images in the fracture group and 5.552 images in the no fracture group. This classification was checked and verified by a radiology registrar with 3 years' radiology experience. The network used from this purpose was InceptionV3 [15] originally trained with ImageNet [16] dataset and then adapted and re-trained for the broken/unbroken classification, modifying the top-layer of the network. In this case, the authors demonstrates how transfer learning could be applied with a dataset of images completely non relatable, i.e. ImageNet.

At contrary, Lindsey et al. [17] pre-trained their network with a large dataset of bones images. The aim of their work was to implement a tool that could help doctors in diagnosis, in order to distinguish if a wrist bone is fractured or not and which part of the bone is fractured. The dataset consisted of 135.845 radiographs of a variety of body parts. Of these, 34.990 radiographs (Training Set) were posterior–anterior or lateral wrist views. The remaining 100.855 radiographs (Pre-Training Set) belonged to 11 other body parts: foot, elbow, shoulder, knee, spine, femur, ankle, humerus, pelvis, hip, and tibia. Every train image was labelled with a bounding box drawn by a group of senior orthopedic surgeons specialized in fractures. The model was a deep CNN, whose architecture is an extension of the common U-Net [18] model. The CNN has two outputs: the probability that the radiograph has a visible fracture and a heat map indicating for each location in the image the probability that the fracture is present in that location. The training of the model can be divided into two stages. In the first stage the model was pre-trained on the Pre-Training Set. In the second stage, the obtained model obtained was fine-tuned using the Training Set, to specialize it to the task of detecting and localizing wrist fractures. After the training and testing phases of the CNN, the authors ran a controlled experiment with 40 emergency medicine clinicians, to evaluated each clinician's ability to detect fractures in wrist radiographs both with and without the help of the system. With the use of the proposed system, the clinicians average sensitivities and specificity improved. This study showed that specialists evaluation may be improved with the use of this system. This procedure should be applied in each work aiming at showing that a CAD system could help humans in evaluation. Pre-training the model before training it with the wrist bones images seem like a good procedure to adjust the parameters for the task, instead of using weights taken from a network that has been trained with a completely different dataset, e.g. ImageNet.

Unfortunately, not always you can count on large dataset for pre-train your network: Rajpurkar et al. [19] solved this problem with the introduction of MURA, one of the largest public radiographic images datasets. MURA dataset contains 14.863 musculoskeletal studies of the upper extremity. Each study contains one or more images taken from different views, with a total of 40.561 images, and was manually labelled by radiologists as normal or abnormal. The studies are divided in 9.045 normal and 5.818 abnormal for 7 different extremities including the shoulder, humerus, elbow, forearm, wrist, hand, and finger. The total number of multi-view images is 40.561. The dataset used by the authors is freely available, which implies that may be used by other researchers to pre-train a model for bones classification that can rely only on a small dataset.

Another good practices is to test different networks before choosing the one that gave the best performance, for example in the work of Olczak et al. [20]. The aim of this study was to assess if standard Deep Learning networks can be trained to identify if a bone is fractured or not in orthopedic radiographs. The dataset was composed of 256.458 hand, wrist, and ankle radiographs, with associated radiologist reports. The authors selected 5 common deep networks for this task: BVLC Reference CaffeNet network (8 layers), VGG CNN S network (8 layers), VGG CNN (16 and 19 layers' networks) and Network-in-network (14 layers). The networks were pre-trained on the ImageNet dataset and the last fully connected layer was replaced in order adapt the network for this specific task. As VGG16 had the best performance in the fracture class, the authors selected it for manual review. When comparing the network with the two senior orthopedic surgeons they found that the network performed similarly as the humans. Testing different existing networks and chose the one that performs best is a good practice in the field of neural network. The dataset contained a really high number of images, the highest among the datasets used in the papers we reviewed. This is obviously one of the most fundamental aspect when working with Deep Learning. With a huge dataset is not easy to label the images manually, and that is why the authors decided to label them automatically from the hospital information. This procedure is subject to errors and a second review may be useful.

Since now we discussed works which obtain good results working with large dataset: Yahalomi et al. [21] proposed a new technique to avoid this. The authors trained a Faster R-CNN [22], a machine vision neural network for object detection, to identify and locate distal radius fractures.

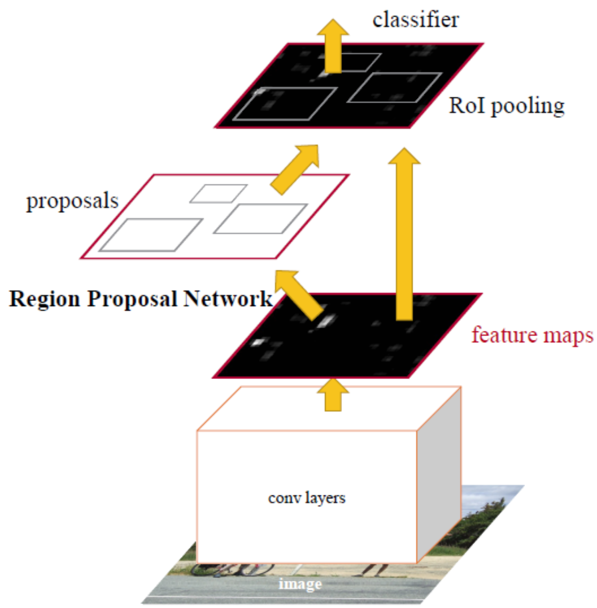


Figure 4. Faster-RCNN architecture [22].

The initial dataset was composed by just 55 images of distal radius fractures and 40 images of hands without fractures. In addition, 25 images not showing hand bones are used for the negative test set. Each image was labelled with bounding boxes around the fractured area. The authors used a Faster R-CNN to achieve two different tasks: classifying weather the fracture is present or not and finding the fracture's location. Faster R-CNN is an evolution of R-CNN and Fast R-CNN where region proposals are generated by CNNs rather than using selective search. Faster R-CNN have 3 different phases. At first, the input images go through a CNN that extract feature maps. Secondly, a RPN (Region Proposal Network) is used for generating region proposal i.e. to pre-check which location contains an object without classifying the entity of the object. The output is then passed through a ROI (Region Of Interest) pooling to perform max pooling on inputs of non-uniform sizes and obtain fixed-size feature maps. Finally, the pooled area goes through CNN and two fully connected branches for class softmax and bounding box regressor, in order to detect the object class and returning the bounding box of that object. The neural network used in this work was VGG16. This is the only work we found that implements the technology of R-CNN to not only classify fracture but also detect the exact region of the fracture with a high accuracy (the results are demonstrated to be significantly more accurate than the detection achieved by physicians and radiologists).

Another improved that can be applied is to select some specific region of the images to improve the network performance, as in Thurston et al. [23]. The aim of this paper was to improve the performance of the system described and already discussed from Kim & MacKinnon [14]. The improvement of this work is given by removing unnecessary parts of the image with semi-automated cropping process. The region of interest was defined using the Python OpenCV [24] *matchTemplate()* function. The method takes a template image and slides it across every position in the subject image (the wrist radiograph), returning the position in which the closest match was calculated. In this study, the region of interest was the distal radius. The template was, therefore, an anatomical representation of the distal radius. The template was produced by using a representative lateral wrist radiograph and applying a smoothing algorithm followed by a binary threshold to segment the bone. The process is showed in Figure 5.

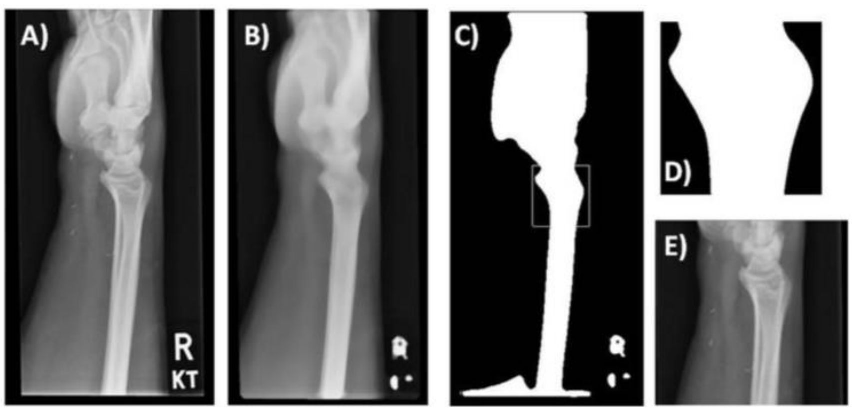


Figure 5. Pre-processing phases in [25].

A scaled template matching approach was adopted to account for different wrist sizes. The accuracy of the model was improved when the region of interest focusing was applied. This extension study has demonstrated that the accuracy of the network to predict fractures can be increased by removing surplus imaging data. The process is still semi-automated, it should be possible to make it fully automated using different techniques, for example using feature matching that tries to match features between the template and the image.

Neural networks are often defined as “black box”, as it’s really hard to understand the learning process. This can be avoided with some visualization techniques, such as the one used by Cheng et al. [26]. The aim of this work is to use a CNN to classify and localize hip fractures on plain frontal pelvic radiographs. The localization phase is implemented by the use of gradient-weighted class activation mapping or Grad-CAM [27] to confirm the validity of the model. The authors used DenseNet [28] network for the classification task. To demonstrate that the CNN is actually focusing on the right area of the images, the authors implemented Grad-CAM to generate a heat map (Figure 6) in the images that the network classified as fractured. The heat maps computed with Grad-CAM were reviewed: after analyzing 49 heat map images, only two images identified the wrong activation site. The use of Grad-CAM can confirm that the network is actually focusing on the correct area of images.

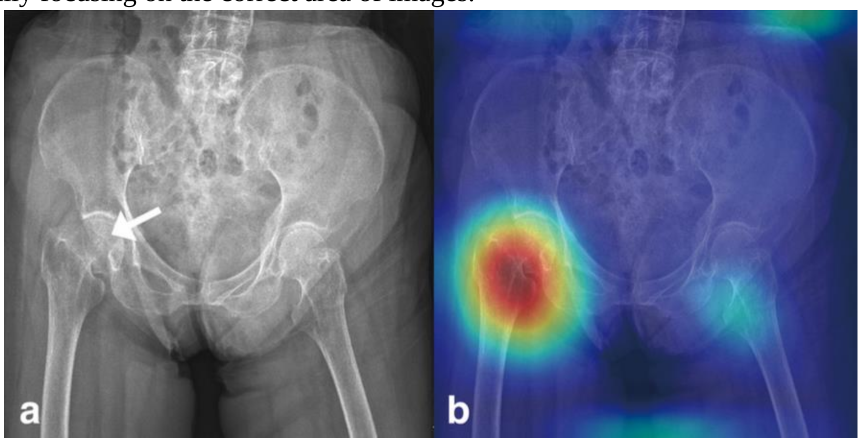


Figure 6. Heatmap generated by Grad-CAM [26].

Since now, we just dealt with paper tackling with binary classification between broken and unbroken bones. The first work concerning the classification in different fractures was proposed by Chung et al. [29]. This work addressed the problem of classification in different types of fractures in the proximal humerus bone. To evaluate the performance of fracture classification, the authors refers to Neer's classification, which is the most commonly used classification for the proximal humerus fracture and distinguish 4 different types of fracture: greater tuberosity (B), surgical neck (C), 3-part (D), and 4-part (E) (Figure 7).

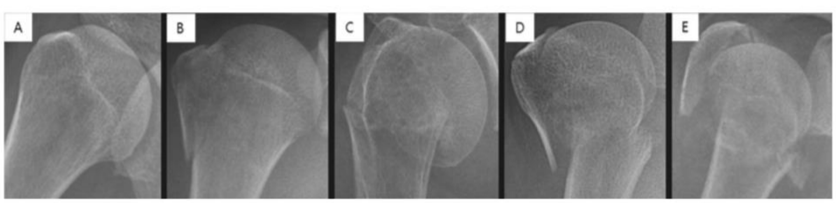

Figure 7. Neer's classification for proximal humerus [29].

Healthy humerus group is named A. Fracture classification was performed by two shoulder orthopedic specialists with 14 and 17 years of experience and one radiologist with expertise in musculoskeletal diseases and 15 years of experience. 515 cases were labelled as A, 346 cases as B, 514 cases as C, 269 cases as D, and 247 cases as E. The dataset of the 1.891 images was divided into 10 partitions without overlapping images: 1 partition was used as a test dataset, while all other images were used as training datasets. The authors used the open source pre-trained ResNet-152 [30] as a deep CNN model. As the dataset was divided into 10 partitions, 10 experiments were performed in order to obtain an averaged performance. The ResNet-152 showed superior performance to that of general physicians and general orthopedists and similar performance to that of the shoulder orthopedists. The authors also used 10-fold cross validation, showed in Figure 8, which is a good practice to reduce evaluation biases. The labelling phase should include different specialists, as this level of fracture is really complex to classify.

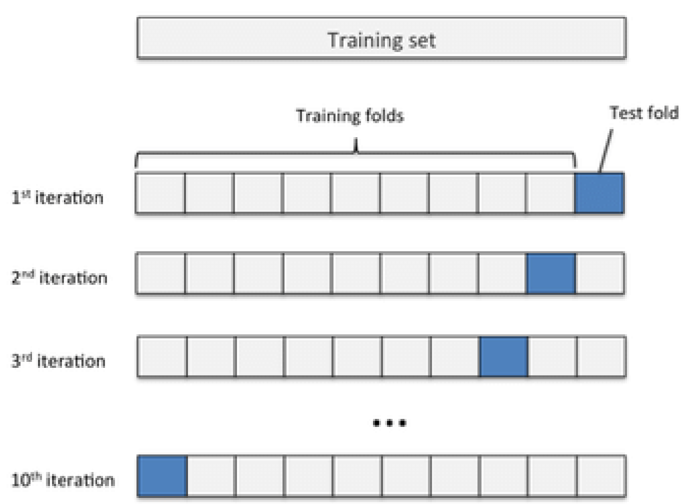


Figure 8. K-fold cross validation graphical representation. This process consist in dividing the dataset in K folds. For each iteration the network uses K-1 folds for training and the remaing fold for testing. In this way it is possible to obtain more generalized results.

A limit of this work is the use of a specific classification for the proximal humerus fracture: to define a generalized approach, we should use a classification structure that can be applied to different bones. In the work of Jiménez-Sánchez et al. [31], the authors proposed a fully automatic CAD tool able to identify, localize and finally classify proximal femur fractures on X-Rays images according to the AO classification. Following AO classification, proximal femur fractures are divided into three main groups: A, B and C, depending from the area that is involved. Each of these classes are subsequently divided into sub-groups. The dataset is composed by a total of 1.347 X-Ray images. For the two classes problem 780 fracture images and 567 normal images were considered. The same dataset was used for the three class problem considering 327 images of type A fractures, 453 of type B fractures and 567 normal X-Rays. Three clinical experts participated in the evaluation: one 5th-year resident trauma surgeon, one trauma surgery attendant and one senior radiologist. Using AO foundation classification is a perfect approach to define a generalized method. The main obstacle broached by the authors is the large imbalance in the frequency of appearance of the classes in the fine-grained classification. A method to deal with unbalanced data must be defined.

## 5. Discussion and Conclusion

Most of the work focused on classification between broken and unbroken bones, without extending the task to different types of fractures. We think that a generalized tool, able to distinguish different types of fractures, should follow the classification stated by the AO foundation. The AO classification is hierarchical, and is determined by the localization and configurations of the fracture lines, where each bone is divided in subsequent sub-groups of fracture, as shown in Figure 9 for the case of proximal femur. In the common literature, the AO classification was claimed to present a better reproducibility compared to other classification systems [32] and its configuration made it optimal for a classification task. Plus, the structure is the same for different bones in the human body, so the approach could be easily extended. Once defined the correct classification system, an adequate dataset is certainly one of the most important aspect for a Deep Learning based application to operate efficiently. Even if in some work good results have been obtained for the fracture/no fracture classification without using a large dataset [13,21] a correct number of images is suggested when the network has to

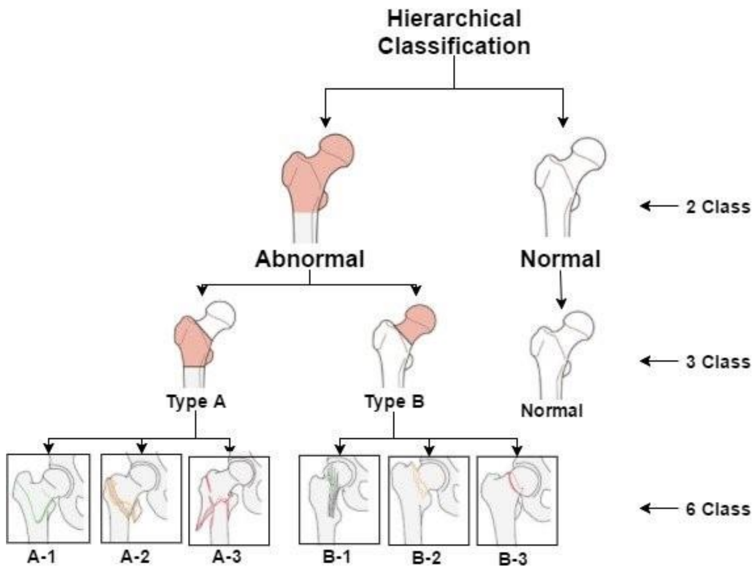


Figure 9. AO classification for the proximal femur case [31].

distinguish different sub-groups of fractures. The dataset could be increased and balanced with data augmentation techniques, if needed, but without adding useless or misleading information. For example, using shear, strain or spot noise augmentation could cause a normal bone image to be classified as a bone with a fracture [21]. Thus, data augmentation is not always enough to balance irregular datasets. One idea to tackle this could be assign different weights to different classes when computing the cost function - classes with few images will be associated with higher weights. Also, technologies such as Generative Adversarial Networks [33] might be used to generate “fake” fractured bones, but this could probably be unfeasible as the fractures might result unrealistic. Concerning the pre-processing phase, the dataset should be cleaned from images containing prosthesis or other evident defects, and we recommend using no more than one image per person to decrease the over performance by the inclusion of a very similar image of the same patient. It is also demonstrated that selecting the fractured area and feed the neural network with cropped images instead of the full image improves the network performances [25,31]. For this reason, a fully automated tool to select the fractured regions should be designed for the pre-processing phase. In addition, is suggested to use lossless format such as PNG and TIFF and to resize images to different sizes and see which one works best [21]. Unfortunately, this is not feasible when using an existing network with transfer learning, because the input images must have a fixed size. Transfer learning allows to use a network, pre-trained on a different dataset, for your own dataset.

The most used network architectures were VGG, ResNet, DenseNet and Inception, pre-trained with ImageNet dataset. As demonstrated in different papers, pre-training the network using a larger dataset of X-Ray bones images may improve the performance [17,19]. For example, MURA dataset [19] is one of the biggest bone's dataset freely available. For this reason, we recommend to try different networks pre-trained with MURA dataset and test which one works best for the specific problem at hand. Cross-validation should be used to demonstrate that the network correctly generalizes the dataset features [29]. Another improvement could be introduced by removing surplus imaging data [29]. For example, if a network has to classify between the no fracture class and three different types of fractures A, B and C, it works as its best if trained excluding no fracture images. Following this results and the AO foundation classification, one idea should be to apply a hierarchical approach. To be more clear, a first network that classify between fracture and no fracture and a sub-sequent one that takes the images predicted as fracture and classify them in A, B and C. Finally, Class Activation Mapping or similar technologies should be used to see where the network is focusing [19,26,31]. Last but not least, also the specialists have a fundamental role: the dataset must be correctly labelled, and different years of experience are needed to properly classify the types of fractures following the AO classification, especially the subgroups. Both for labelling and evaluation, if possible, more than one expert, coming from different specialization, should be enrolled in order to have multiple opinions. As the final aim of this tool would be to prove that the CAD system effectively help doctors in diagnosis, it would be also important to evaluate the performance of the specialists with and without the help of it [17]. A summary of the main weaknesses and strenghts of each paper is showed in Tab. 1.

In this section we outlined the main aspects that should be taken in consideration when trying to tackle this task. This will be a first step in order to develop a solution that includes all these aspects and could be applied to classify all the bones in the human body, bringing a huge progress in the orthopaedic field.

| | Main weakness | Main strenght |
|---|---|---|
| Dimililer et al. [13] | Scant and bad-structured dataset | Pre-processing phase |
| Kim and MacKinnon [14] | Network pre-trained with a non-relatable dataset | Used transfer learning for the first time in this sub-field |
| Lindsey et al. [17] | | Very large dataset for pre-training and training |
| Rajpurkar et al. [19] | Low results compared to the dimension of the dataset | The dataset used was made freely available by the authors |
| Olczak et al. [20] | Images labelled automatically without a second review | Used different networks structures |
| Yahalomi et al. [21] | Labelling carried on by just one specialist | Optimal results with a small dataset |
| Thurston et al. [25] | The region detecton phase is semi-automated | Improved performance with the use of specific region |
| Cheng et al. [26] | | Used CAM to visualize the learning process |
| Chung et al. [29] | Used a non-extendable classification method | Classification in different type of fractures |
| Jimenez-Sanchez et al. [31] | Imbalanced dataset | Used AO classification |

Tab.1 Main strengths and weaknesses of each paper. Two papers dos not present any main weakness: this does not mean that they do not have weaknesses, but that their weaknesses are due to some missing aspects and not to some wrong implementations.

## 6. REFERENCES


[1] Amparore D, Checcucci E, Gribaudo M, Piazzolla P, Porpiglia F, Vezzetti E. Non-linear-Optimization Using SQP for 3D Deformable Prostate Model Pose Estimation in Minimally Invasive Surgery. In: Arai K, Kapoor S, editors. Adv. Comput. Vis., vol. 943, Cham: Springer International Publishing; 2020, p. 477–96. https://doi.org/10.1007/978-3-030-17795-9_35.

[2] Olivetti EC, Nicotera S, Marcolin F, Vezzetti E, Sotong JPA, Zavattero E, et al. 3D Soft-Tissue Prediction Methodologies for Orthognathic Surgery—A Literature Review. Appl Sci 2019;9:4550. https://doi.org/10.3390/app9214550.

[3] International Osteoporosis Foundation. Broken bones, broken lives: a roadmap to solve the fragility fracture crisis in Europe 2018. https://www.iofbonehealth.org/broken-bones-broken-lives (accessed September 20, 2019).

[4] Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Kim J. Long Radiology Workdays Reduce Detection and Accommodation Accuracy. J Am Coll Radiol 2010;7:698–704. https://doi.org/10.1016/j.jacr.2010.03.004.

[5] Hallas P, Ellingsen T. Errors in fracture diagnoses in the emergency department – characteristics of patients and diurnal variation. BMC Emerg Med 2006;6:4. https://doi.org/10.1186/1471-227X-6-4.

[6] Checcucci E, Autorino R, Cacciamani GE, Amparore D, Cillis SD, Piana A, et al. Artificial intelligence and neural networks in urology: current clinical applications. Minerva Urol Nefrol 2019. https://doi.org/10.23736/S0393-2249.19.03613-0.

[7] Cirrincione G, Marcolin F, Spada S, Vezzetti E. Intelligent Quality Assessment of Geometrical Features for 3D Face Recognition. In: Esposito A, Faundez-Zanuy M, Morabito FC, Pasero E, editors. Neural Adv. Process. Nonlinear Dyn. Signals, vol. 102, Cham: Springer International Publishing; 2019, p. 153–64. https://doi.org/10.1007/978-3-319-95098-3_14.

[8] Kumar R. Research Methodology: A Step-by-Step Guide for Beginners. SAGE; 2019.

[9] Al-Ayyoub M, Hmeidi I, Rababah H. Detecting hand bone fractures in x-ray images. JMPT 2013;4:155–168.

[10] Cao Y, Wang H, Moradi M, Prasanna P, Syeda-Mahmood TF. Fracture detection in x-ray images through stacked random forests feature fusion. 2015 IEEE 12th Int. Symp. Biomed. Imaging ISBI, Brooklyn, NY, USA: IEEE; 2015, p. 801–5. https://doi.org/10.1109/ISBI.2015.7163993.

[11] Myint WW, Tun HM, Tun KS. Analysis on Detecting of Leg Bone Fracture from X-ray Images. Int J Sci Res Publ IJSRP 2018;8. https://doi.org/10.29322/IJSRP.8.9.2018.p8150.

[12] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44. https://doi.org/10.1038/nature14539.

[13] Dimililer K. IBFDS: Intelligent bone fracture detection system. Procedia Comput Sci 2017;120:260–7. https://doi.org/10.1016/j.procs.2017.11.237.

[14] Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol 2018;73:439–45. https://doi.org/10.1016/j.crad.2017.11.015.

[15] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, Las Vegas, NV, USA: IEEE; 2016, p. 2818–26. https://doi.org/10.1109/CVPR.2016.308.

[16] Fei-Fei L, Deng J, Li K. ImageNet: Constructing a large-scale image database. J Vis 2010;9:1037–1037. https://doi.org/10.1167/9.8.1037.

[17] Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, et al. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci 2018;115:11591–6. https://doi.org/10.1073/pnas.1806905115.

[18] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Med. Image Comput. Comput.-Assist. Interv. – MICCAI 2015, vol. 9351, Cham: Springer International Publishing; 2015, p. 234–41. https://doi.org/10.1007/978-3-319-24574-4_28.

[19] Rajpurkar P, Irvin J, Bagul A, Ding D, Duan T, Mehta H, et al. MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs. ArXiv171206957 Phys 2018.

[20] Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs: Deep learning algorithms—are they on par with humans for diagnosing fractures? Acta Orthop 2017;88:581–6. https://doi.org/10.1080/17453674.2017.1344459.

[21] Yahalomi E, Chernofsky M, Werman M. Detection of Distal Radius Fractures Trained by a Small Set of X-Ray Images and Faster R-CNN. In: Arai K, Bhatia R, Kapoor S, editors. Intell. Comput., vol. 997, Cham: Springer International Publishing; 2019, p. 971–81. https://doi.org/10.1007/978-3-030-22871-2_69.

[22] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell 2017;39:1137–49. https://doi.org/10.1109/TPAMI.2016.2577031.

[23] Thurston M, MacKinnon T, Kim DH. Fracture detection with artificial intelligence: improved accuracy with region of interest focusing, 2018.

[24] Bradski G. The OpenCV Library. Dr Dobb's Journal of Software Tools; 2000.
[25] Thurston M, MacKinnon T, Kim DH. Fracture detection with artificial intelligence: improved accuracy with region of interest focusing, 2018.
[26] Cheng C-T, Ho T-Y, Lee T-Y, Chang C-C, Chou C-C, Chen C-C, et al. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. Eur Radiol 2019;29:5469–77. https://doi.org/10.1007/s00330-019-06167-y.
[27] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE Int. Conf. Comput. Vis. ICCV, Venice: IEEE; 2017, p. 618–26. https://doi.org/10.1109/ICCV.2017.74.
[28] Huang G, Liu Z, Maaten L van der, Weinberger KQ. Densely Connected Convolutional Networks. 2017 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, Honolulu, HI: IEEE; 2017, p. 2261–9. https://doi.org/10.1109/CVPR.2017.243.
[29] Chung SW, Han SS, Lee JW, Oh K-S, Kim NR, Yoon JP, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop 2018;89:468–73. https://doi.org/10.1080/17453674.2018.1453714.
[30] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, Las Vegas, NV, USA: IEEE; 2016, p. 770–8. https://doi.org/10.1109/CVPR.2016.90.
[31] Jiménez-Sánchez A, Kazi A, Albarqouni S, Kirchhoff C, Biberthaler P, Navab N, et al. Towards an Interactive and Interpretable CAD System to Support Proximal Femur Fracture Classification. ArXiv190201338 Cs 2019.
[32] Jin W-J, Dai L-Y, Cui Y-M, Zhou Q, Jiang L-S, Lu H. Reliability of classification systems for intertrochanteric fractures of the proximal femur in experienced orthopaedic surgeons. Injury 2005;36:858–61. https://doi.org/10.1016/j.injury.2005.02.005.
[33] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks. ArXiv14062661 Cs Stat 2014.