

## A Systematic Review on Artificial Intelligence in Orthopedic Surgery

Nabila Ounasser<sup>1\*</sup>, Maryem Rhanoui<sup>1,2</sup>, Mounia Mikram<sup>2</sup>, Bouchra El Asri<sup>1</sup>

<sup>1</sup> ADMIR Laboratory, ENSIAS, Mohammed V University, Rabat 10112, Morocco

<sup>2</sup> LYRICA Laboratory, School of Information Sciences, Rabat 10112, Morocco

Corresponding Author Email: [nabilaounasser81@gmail.com](mailto:nabilaounasser81@gmail.com)

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ria.380409>

### ABSTRACT

**Received:** 12 December 2023

**Revised:** 2 March 2024

**Accepted:** 12 April 2024

**Available online:** 23 August 2024

#### Keywords:

*Artificial Intelligence, Deep Learning, machine learning, Generative Adversarial Network, Convolutional Neural Network, orthopedic, anomaly diagnosis, medical image*

This systematic review aims to assess the efficacy of Artificial Intelligence (AI) applications in orthopedic surgery, with a focus on diagnostic accuracy and outcome prediction. In this review, we expose the findings of a systematic literature review awning the papers published from 2016 to October 2023 where authors worked on the application of an AI techniques and methods to an orthopedic purpose or problem. After application of inclusion and exclusion criteria on the extracted papers from PubMed and Google Scholar databases, 75 studies were included in this review. We examined, screened, and analyzed their content according to PRISMA guidelines. We also extracted data about the study design, the datasets included in the experiment, the reported performance measures and the results obtained. In this report, we will share the results of our survey by outlining the key machine and Deep Learning (DL) techniques, such as Convolutional Neural Network (CNN), Autoencoders and Generative Adversarial Network, that were mentioned, the various application domains in orthopedics, the type of source data and its modality, as well as the overall quality of their predictive capabilities. We aim to describe the content of the articles in detail and provide insights into the most notable trends and patterns observed in the survey data.

## 1. INTRODUCTION

The healthcare industry has experienced a notable upswing in attention towards Artificial Intelligence (AI), which has been steadily transforming medical procedures. With the advancements in data storage and computer processing power, computer systems are acquiring the capacity to accomplish tasks that previously necessitated human intelligence.

Historically, AI in orthopedic surgery has evolved significantly over the decades, from early rule-based systems in the 1950s to sophisticated Machine Learning (ML) algorithms and robotic-assisted surgery today. Pioneering efforts like MYCIN in the 1970s and computer-assisted orthopedic surgery (CAOS) systems in the 1990s demonstrated AI's potential in medical decision-making and enhancing surgical precision. In the 2000s, the integration of ML enabled predictive modeling and personalized treatment planning, while the adoption of robotic systems in the 2010s revolutionized surgical techniques. Today, AI applications in orthopedics focus on leveraging patient-specific data and predictive analytics to optimize outcomes and tailor treatments, promising computer-aided detection (CAD) in musculoskeletal healthcare.

CAD systems and Diagnostic Imaging are pivotal in orthopedic disease diagnosis, leveraging AI to enhance clinical decision-making. By analyzing medical images, such as X-ray, MRI, and CT scans, clinicians can visualize bone structures and soft tissues, crucial for identifying injuries and

conditions. However, interpreting these images can be intricate and time-consuming. Therefore, developing a CAD system for anomalies detection has become attractive in medical imaging. Enter AI, which has revolutionized orthopedic imaging by swiftly and accurately analyzing medical images. In fact, it has become the preferred approach for analyzing radiology images. This includes different tasks like bone tumor detection, cartilage segmentation, spinal disease prediction etc. on different image modalities Xray, MRI, CT scans, etc. The potential for improvement in patient care through these means is broad, encompassing areas such as diagnosis, management, research, and systems analysis.

Within this frame of reference, recently, Deep Learning (DL) has made significant advances. It demonstrates the capacity of diagnosing tasks for medical imaging similarly to the performance of human radiology experts, mainly through the utilization of special and powerful architectures as Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN). DL algorithms have shown remarkable proficiency in executing various radiographic tasks in the field of musculoskeletal radiology, exhibiting an expertise level that enables accurate diagnosis of orthopedic diseases. Several papers have been worked on detecting bone anomalies by exploring DL models, with good to excellent accuracy even similar to expert human performance in a minimal time with much faster speeds. Although these findings are interesting, no systematic study has been conducted to analyze the scope and efficacy of AI algorithms,

especially GAN-based architecture in orthopedic anomalies diagnosis.

The scope of AI applications within orthopedic surgery covered in this review encompasses the utilization of AI techniques, particularly focusing on DL models such as GANs, AEs and CNNs, in various diagnostic tasks related to orthopedic conditions. Specifically, the review addresses the application of AI in bone diagnosis, musculoskeletal imaging, fracture detection, spine pathology diagnosis, and cartilage diagnosis. Furthermore, the review evaluates primary research studies that have developed AI algorithms for diagnosing orthopedic diseases, including bone tumors detection, fracture detection, spinal pathology diagnosis, and cartilage diseases. The emphasis is on summarizing the methods used for training AI algorithms, managing datasets, and assessing the accuracy of these algorithms in diagnosing orthopedic pathologies through medical imaging. By delineating this scope, the review aims to provide readers with a focused understanding of the current state and potential applications of AI in orthopedic surgery.

This review focuses on AI applications that have the potential to enhance or transform clinical practice. It provides a historical overview of AI in medicine to contextualize recent developments, highlights successful application areas, and identifies potential avenues for future research. AI is still a

novel concept for many orthopedic surgeons, but the existing body of work provides valuable insights into potential new applications and areas for research. In this systematic review, we examine the use of CNN and GAN-based architectures in the literature. We will thoroughly discuss the current state of these models and the results of the included studies to understand the extent of AI research in orthopedics, describe how AI has been applied in the field, and provide a glimpse into its potential future applications.

## 2. MATERIALS AND METHODS

To conduct a thorough investigation of AI articles regarding orthopedic disease diagnosis, we conducted a literature search using PubMed and Google Scholar. The key-words used for both medical and AI components are listed in Table 1. The articles had to include at least one keyword from the medical or AI part in either the title or abstract. We included the studies published since 2016 up to October 2023, this search time frame was chosen according the timeline in the Figure 1 that we investigated to ensure having export papers related to our purpose. Then we had scanned articles according to PRISMA guidelines as shown on Figure 2.

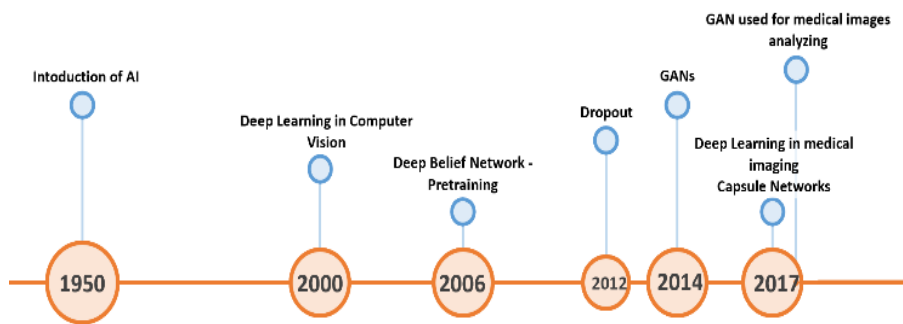


Figure 1. Timeline of work related to Deep Learning and GANs in medical imaging

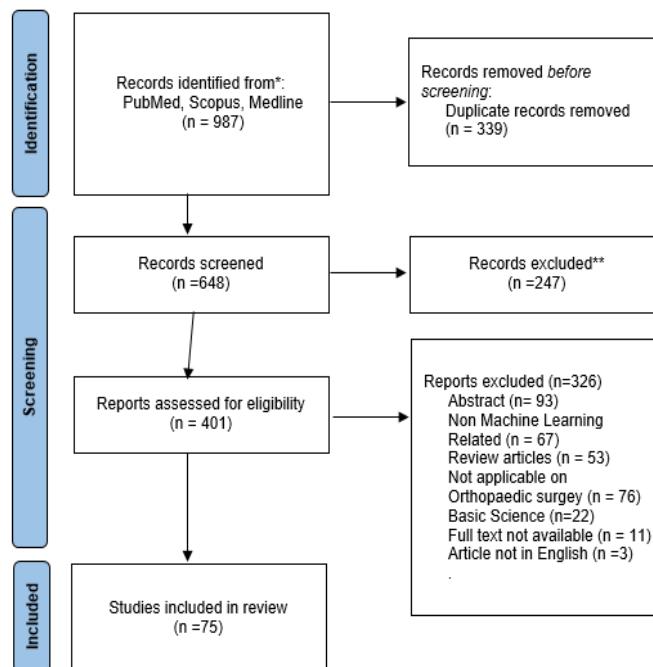


Figure 2. PRISMA flowchart showing systematic review search strategy

**Table 1.** A summary of keywords utilized in the PubMed research and GoogleScholar

Medical Keywords	AI Keywords
Medecine	Artificial Intelligence
Healthcare	Deep Learning
Orthopedics	Generative Adversarial Network
Bone Diagnosis	Convolutional Neural Network
Medical Image	Computer aided diagnosis
Musculoskeletal	Anomaly Detection
Fracture Detection	
Bone Tumors	
Spine Pathology	
Cartilage Diagnosis	
Radiographs	

Notes: The medical and AI keywords were linked using a logical OR operator, while the two groups of keywords were connected using a logical AND.

## 2.1 Inclusion and exclusion criteria

The study focused on summarizing the application of AI in orthopedic diseases diagnosis. After duplicates were removed, the titles and abstracts of each study were screened, and those deemed potentially significant during the screening process were then subjected to a complete examination in their entirety. Hence, we investigated the following inclusion criteria:

- Studies which directly applied AI to diagnosis orthopedic diseases or related and treating at least one of the anomalies involved (fracture, deformation, spine pathology, cartilage anomalies, osteoarthritis).
- Data: The selected articles employed AI methods in the field of computer vision and involved the use of medical images, text data, or clinical data.
- Subjects included in the study: all the papers have to be based on studies of human bones or cartilage and related pathology.
- Aim of the study: An alternate medical issue was excluded from the scope of the study: articles not related to orthopedic diseases and their associated medical data were disregarded. For instance, studies solely focused on COVID-19 detection or oncology pathology diagnosis were not considered.
- GAN and CNN were not considered: we excluded studies that did not implement CNN and GAN-based methods in the diagnosis of different orthopedics anomalies.
- Results: we excluded articles that did not give a performance metrics procedure or reported a clear result of the investigated algorithms.
- Validation procedures: results had to be reported on a test set separate from the training set.
- Review article: we excluded all literature review studies.
- Full text not available: we excluded articles that we could not find the full text to explore.
- Language: Articles written in languages other than English were excluded.
- Animal studies: studies focusing on orthopedic diseases in animals were disregarded.

## 2.2 Data extraction and collation

To conduct a thorough investigation of AI articles regarding orthopedic disease diagnosis, we conducted a literature search using PubMed and Google Scholar. The keywords used for both medical and AI components are listed in Table 1. The articles had to include at least one keyword from the medical

or AI part in either the title or abstract. We included the studies published since 2016 up to October 2023, this search timeframe was chosen according to the timeline in the Figure 1 that we investigated to ensure having export papers related.

## 3. RESULTS

The search was conducted in October 2023 and produced 987 articles. Once the duplicates had been eliminated and a preliminary assessment based on the titles and abstracts had been completed, the total number of qualifying articles was reduced to 75. A second screening phase was carried out after reading the full text. The selection process of the studies was documented using a flowchart diagram following the PRISMA protocol Figure 2. The articles were evaluated for inclusion or exclusion. It was noted that the number of published papers has been increasing annually, with the number of scientific papers published in 2023 being almost double that of 2019. This trend may be attributed to several factors, particularly the growing availability of medical images and data to researchers. In addition, the capacity of Generative models noticed in recent years in anomaly detection tasks. Moreover, medical stuff understands the advantage and the need AI systems to improve clinical performance in term of time consuming in diagnosis and precision.

In this section, we have come up with an overview of the main papers we collected during our literature review exploring AI methods notably CNN and GAN-based architectures [1].

### 3.1 Spine pathologies detection

In this section, we begin to report on papers in which authors have diagnosed spinal diseases using AI methods.

Among the studies that sought to assess many researchers have trained CNN on radiographic images [2-5] Particularly, in the study [6], they developed and validated DL algorithms (DLAs) to automatically detect scoliosis using unclothed back images. Their architecture contains Faster-RCNN that localized the regions of interest (from neck to hip). To identify the characteristics of each group, a resnet of 101 levels was created. After undergoing preprocessing, the Resnet was utilized to extract advanced characteristics for the purpose of binary classifications. The algorithms' accuracy allowed for the detection of cases with a curve of 20° and severity grading for both binary and four-class categories. In the study [7], authors reported an automated approach that can handle spinal abnormalities by collecting anatomical parameters from

biplanar X-rays of the spine. In order to anticipate the position of each landmark, a CNN was completely trained with an added DSNT layer that was spatial to numerical and differentiable. The predictions made by the models regarding the shape of the spine were closely linked to the actual shape. The standard errors of the computed parameters varied from  $2.7^\circ$  (for the pelvic tilt) to  $11.5^\circ$  (for then L1-L5 lordosis). The ground truth shape and the spinal form predicted by the models had a good correlation. CNN models have also been investigated to identify and segment discs and vertebrae from spinal MRIs [8-10]. Notably, He et al. [11] suggest a detection framework SpineOne to locate and categorize degenerating discs and vertebrae from MRI slices. The following three essential methods are the foundation of SpineOne: Three innovations were made: 1) a new keypoint heatmap design to enable simultaneous keypoint localization and classification; 2) Attention modules were explored to distinguish between vertebrae and discs in representations; and 3) A new method, called SpineNet, has been introduced in the paper "SpineNet: Learning Scale-Permuted Backbone for Recognition" [12]. It uses a gradient-based approach to associate multiple learning objectives during later stages of training. SpineNet is based on CNNs and has a unique backbone structure with rearranged intermediate features and connections between different scales. It was trained for object detection through a process called Neural Architecture Search.

Imaging analysis approaches are being introduced potentially for scientific research and clinical use to facilitate spinal pathologies diagnosis. Machine learning-based approaches had been investigated to detect, classify or segment anatomical spinal landmarks [13-16]. Wu et al. [13] introduce a new framework for estimating landmarks in X-ray images for AIS (Automatic Intervertebral Segmentation) assessment. The framework leverages BoostNet, a creative integration of ConvNet and statistical methods. BoostNet has strong feature extraction capabilities and is able to handle the variability in X-ray images. The BoostNet architecture shows a robust quality scoliosis detection in the clinical scenario by estimating a mean squared error (MSE) rate of 0.00068 in 431 cross validation images and 0.0046 in 50 test images. Also, landmarks detection and alignment analysis have been recently treated in whole-spine lateral radiographs by Yeh et al. [14]. The authors [14] propose a DL approach for identifying spinal anatomical landmarks and predicting radiographic parameters. Their approach resulted in predictions that had a high correlation with ground truth values, with all p-values less than 0.001.

Otherwise, statistical methods have been investigated to treat spinal deformities by calculating the full geometry or specific geometrical of spine parameters, such as sagittal vertical axis, lumbar lordosis, and Cobb angle [15-18]. Zhang et al. [15] develop a computer-aided technique using a deep neural network (DNN) that was trained using vertebral patches taken from radiographs of a spinal model. The vertebral slopes predicted by the DNN were used to automatically determine the Cobb angle of the spinal curve. The mean absolute differences for model radiographs were less than  $3^\circ$ , and the intraclass correlation coefficients were higher than 0.98. This shows that the proposed approach had a good level of reproducibility when it came to measuring model radiography. Yi et al. [17] performed a method which demonstrate its merits in both Cobb angle measurement and landmark detection on low-contrast and ambiguous X-ray images. Moreover, Caesarendra et al. [18] proposes a DL architecture to identify

spinal vertebrae from radiographs to automate the calculating of the Cobb angle that helps to diagnosis the presence of scoliosis and spinal deformities. From the input image, algorithm detect the landmark features, then it calculates the Cobb angle. As a result, the suggested algorithm has a classification accuracy of about 90 percent.

The use of GAN-based predictive models could be especially beneficial in spine surgery due to the complex nature of the procedures and potential high rates of complications in patients who often have multiple health issues. Recently, spinal anomalies detection and spinal shape diagnosis through GAN models have been gaining interest. A number of GAN-based models have been performed for spine-related tasks as well as [19, 20] from either MRIs, CT or X-ray images. Authors [20] introduce a SpineGAN model, which uses an atrous convolution autoencoder module to handle the problem of the high variety and variability of complicated spine structures while maintaining fine-grained structural data. Spine-GAN uses a discriminative network that can rectify anticipated mistakes and global-level contiguity to provide reliable performance and effective generalization. Numerous studies using the MRIs of 253 patients showed that SpineGAN achieves high pixel accuracy of 96.2 percent, demonstrating its effectiveness and potential as a therapeutic tool. CycleGAN is one of the most famous GAN models that attracted a lot of attention through image-to-image translation using unpaired images [21]. CycleGAN model uses a trainable preprocessing pipeline that normalizes the input MRI data exploring a low-capacity fully CNN, then uses FC-ResNets in flow for the vertebral bodies segmentation.

It is suggested that a pseudo-3D CycleGAN architecture be used, along with a cyclic loss function to provide coherence across MRI and CT synthesis. It produced promising results, it showed the power of the pipeline by achieving clinically durable CT scans which can be treated for surgical guidance.

In the research [19], authors proposed and verified a semi-supervised GAN approach for early scoliosis detection using chest X-rays. The proposed model employs a semi-supervised training process where the GAN is initially taught to recognize scoliosis patterns and then performs basic classification to accurately differentiate between normal and scoliotic states. The results show that the negative predictive value (NPV) and positive predictive value (PPV) are 0.856 and 0.950 respectively.

### 3.2 Fracture detection

Fracture is the most accurate orthopedic pathology in the most hospitals. Analyzing medical images to identify bone fractures is time-consuming and requires qualified experts. Therefore, Scientists investigate their time to help doctors to reduce diagnosis timing, and to improve decision precision [22-25].

AI/DL has the potential to assist medical professionals and decision makers in creating effective and cost-efficient treatment plans [26-28]. Numerous studies have shown that incorporating automated tools during the diagnostic phase can improve the accuracy of physician interpretation [29, 30].

Most of studies mainly investigate on accurately detecting musculoskeletal abnormalities CNN models. Authors [31-33] have proposed GnCnnr adopting the principle of normalization, including group normalization, weight normalization and cyclic learning rate planner to improve the model performance measures. When compared to other deep

learning techniques like DenseNet, Inception, and Inception v2 model, the GncNNr model had the highest accuracy of 85 percent. Regnard et al. [34] performed AI models for skeletal lesions detection and localization to compare them with the routine radiological interpretation. After a traumatic pelvic and limb injury, they gathered radiographic examinations and the related radiologists' reports. An AI (BoneView, Gleamer) was used to analyze each exam, and the results were compared to the radiologists' reports. An alternative to using CNN is Transfer Learning, a widely used DL approach in computer aided systems. In the study [35, 36], Abreu Dias and Kim and MacKinnon used the Inception v3 network, retrained on lateral wrist radiographs, to improve fracture detection and classification. The model was trained on 11,112 images after a data augmentation pretraining, starting with 1,389 radiographs (695 "fracture" and 694 "nofracture"). The resulting AUC was 0.954.

Now we will focus on studies that surpass the traditional Convolutional Neural Network to work with generative adversarial and AE networks [37]. The Res-unetGAN network [38], a generative adversarial network-based unsupervised anomaly detection approach, is proposed in this article. To calculate the normal distribution of normal samples, the autoencoder architecture contains ResNet50 and UNet that is employed by the generative network. The discriminator employs a Convolutional Neural Network model that is based on deep separable convolution to construct a gaming process. To achieve the goal of identifying anomalies, a reconstruction error score is calculated based on the quality of reconstruction, and the presence of defects in the sample is evaluated by this reconstruction error score. Following numerous tests on the Mura data set, the Mura defect's detection accuracy outperforms several other models. Davletshina et al. [39] showed how unsupervised techniques trained on radiographic images devoid of anomalies can help clinicians assess. The approach focuses on improvement of diagnostic accuracy and lowers the possibility of overlooking critical areas. Therefore, the use cutting-edge techniques for unsupervised learning to find anomalies and demonstrate how the results of these techniques may be justified.

### 3.3 Osteoarthritis detection and prediction

Experts frequently use manual examination of patient medical images, which are typically gathered in hospitals, to make an osteoarthritis diagnosis. Osteoarthritis is a time-consuming task. Therefore, several studies coverage and concentrate on using image-based DL systems to automatically identify osteoarthritis [40, 41].

Jakaite et al. [42] investigate ML techniques to build a comparative study to deduce ML's capability for osteoarthritis (OA) radiography diagnosis at a primal stage, and the number of patient cases is low. With Kellgen-Lawrence scores ranging from 1 to 10, patients' knees were detected in high-resolution X-ray scans for their investigations. Although the Group Method of Data Handling approach in DL has shown a significant enhancement in diagnostic testing, the current ML methods have only demonstrated a slight increase in diagnostic precision. The comparative trials show that the suggested framework using texture features based on Zernike has greatly increased diagnostic accuracy, increasing it by an average of 11 percent. Wang et al. [43] propose an end-to-end approach to automatic osteoarthritis diagnosis by combining into the training workflow a YOLO object detection algorithm

and a visual transformer. Their strategy requires to analyse 200 annotated images from a huge dataset with more than 4500 samples, but it accurately segments 95.57 percent of the data. Additionally, compared to CNN based models developed for the same case, their classification result increased accuracy by 2.5 percent. It performed patient statistics on medical use and health behavior variables to train a deep neural network (DNN) to detect the presence of osteoarthritis [44]. From the patients' basic background medical records, characteristics were generated using principal component analysis (PCA) with quantile transformer going over to determine the presence of osteoarthritis. Our tests demonstrated that the suggested approach, which combined a deep neural network with scaled PCA, produced an area under the curve of 76.8 percent while requiring the least amount of feature generation work. Therefore, to cut down on medical expenses and the amount of time patients spend in hospitals, this strategy could be a hopeful tool for patients and clinicians to screen in advance for potential osteoarthritis.

### 3.4 Bone and cartilage image diagnosis

Recently, research has explored the role of AI and has come to increasingly recognize the significance of DL in the medical domain, particularly in the area of computer-assisted knee osteoarthritis diagnosis [45, 46]. They highlighted the potential value of diagnostic approaches for early knee osteoarthritis identification namely bone segmentation [47, 48], bone classification [49] and abnormality detection [50, 51]. Most of studies mainly adapted on accurately diagnosis musculoskeletal abnormalities CNN models. For diagnosing abnormal musculoskeletal radiographs, He et al. [52] present an innovative calibrated ensemble of deep learners. Their model makes use of the advantages of most fundamental DL networks (DenseNet, ConvNet and ResNet), which are frequently used straight fully or as the framework in other DL based methods. The introduced model showed perspective results comparing to three individual models and a conventional ensemble learner, according to experimental findings based on the publicly available MURA dataset. Their model achieved an "overall efficiency of (Accuracy: 0.87, Precision: 0.93, AUC: 0.93, Recall: 0.81, Cohen's kappa: 0.74). Using a convolutional models, Noguchi et al. [53] create and assess an algorithm for bone segmentation on whole-body CT. They assessed the effectiveness of the different data augmentation techniques to enhance the network's performance and robustness (RICAP). The network's mean Dice coefficient was 0.983 0.005 after training on the internal dataset. It demonstrated the effectiveness of convolutional based architectures (CNNs) for generalized abnormality detection on radiographs of the lower extremities [54]. They gathered a sizable dataset of 93,455 radiographs of the lower extremities of various body areas, classifying each exam as normal or abnormal. On this abnormality classification test, they had implemented a 161-layer densely connected and pre-trained CNN to attend an AUC-ROC of 0.880 (specificity = 0.961, sensitivity = 0.714).

Lately, researchers rely on GAN models to innovate perspective solutions for medical issues. A number of GAN-based models have been performed for bone and cartilage-related tasks as well as [55, 56] from either MRIs, Ultrasound or X-ray images. Alsilan et al. [57, 58] investigates his time for exploring DL on Bone pathologies diagnosis. suggest a computational technique based on GAN architecture that can

simultaneously generate segmented bone surface masks and synthetic B-mode US images. Alsinan et al. [57] suggested GAN model produced realistic B-mode bone US images and segmented bone masks using two convolutional layers. On 1235 images taken from 27 patients using two distinct US machines, quantitative and qualitative assessment tests are carried out to demonstrate a comparison of their model's results with the current leading GANs for bone segmentation task surfaces using a U-Net architecture. In the study [58], Alsinan suggested a real time computational technique to separate bone shadow pictures from in vivo US scans based on a novel GAN architecture. Additionally, he demonstrates the potential of using the segmented shadow images as a substitute for accurate bone surface segmentation in real-time through the application of a multi-feature supervised CNN architecture. They were able to segment bone shadows with a mean dice coefficient of 93 percent (0.02), demonstrating that the system is comparable to manual expert annotation. In the study [59], authors used GAN and UNet models to segment bone features from a wrist US scan. The ensemble models were implemented on 10,500 wrist US scans from 47 patients obtained from the pediatric emergency department of the University of Alberta Hospital (UAH). In general, GAN had the strongest recall even though UNet had the highest DICE score, accuracy, and Jacard Index.

### 3.5 Skeletal bone age evaluation

Bone age is a constructive indicator used to radio-logically evaluate and diagnosis various bone pathologies and indentify the most suitable medicine and the optimal timing of treatment. The purpose of bone age assessment is to measure growth and maturity and to treat pediatric disorders. Recently, several AI researchers work on developing computer aided systems for bone age assesment. In this section, we review studies that investigate DL based approaches on bone age assesment. Ren et al. [60] reported a DL-based approach for automatically training hand radiographs for a bone age assesment. The network was designed to target bone age-related regions in the images and incorporates an attention module for generating coarse and fine attention maps to be fed into the regression network. Additionally, the regression network is supervised with a dynamic attention loss, allowing it to more accurately estimate the bone age even for challenging or "outlier" images. The results of the experiment show that the proposed approach has an average discrepancy of 5.2-5.3 months between clinical and automatic bone age assessments when tested on two large datasets.

A new approach [61] to improve BAA training throughout the pretraining and training process was introduced using GAN. The pre-training framework uses a unique distance metric called cosine distance, which is applied to optimal transport for data augmentation (CNN-GAN-OTD). During the training phase, a method that combined data. A DL-based computer-aided diagnosis was developed by Li et al. [62] for the purpose of performing bone age assessments. Firstly, during image-processing pipeline they reduced by exploring an unsupervised learning that identify informative regions. Accordingly, to increase the accuracy of prediction, they used a backbone image model with pre-trained parameters. The best outcomes from the experimental comparative study revealed a mean absolute error on the public RSNA dataset of 6.2 and 5.1 months on the supplementary dataset utilizing MobileNetV3 as the foundation.

### 3.6 Bone tumors detection

Cancer is widely recognized as one of the most perilous illnesses globally. In medical terms, it is called a malignant neoplasm. This genetic disorder is brought on by uncontrolled cell growth. Early identification of this dangerous condition can reduce mortality rates [63, 64]. The X-ray images is explored to detect, classify or segment bone cancer [65]. Park et al. [66] developed and validate an Artificial Intelligence-based primal identification and classification of bone tumors in the proximal femur on X-ray images. A single tertiary referral center provided 538 standard anteroposterior hip X-ray images, 94 (120 benign, 94 malignant, or 324 no tumor), were used to train the AI model. The image modalities were pre-trained to make them ideal for the deep learning algorithm's training. CNN models were used to conduct the multi-classification on each femur using pre- processed pictures. As a result, the best CNN model has an AUROC of 0.953. von Schacky et al. [67] and He et al. [68] followed the same strategy by adopting DL approaches to diagnosis primary bone cancers based on clinical radiographs. All patients had their bone tumors classified as benign or malignant using the histopathologic results as the gold standard. The internal data set includes radiographs from 934 individuals, 667 of which were benign bone tumors, and 267 of which were malignant. The multitask DL model classified bone lesions as benign or malignant with an accuracy of 80.2 percent. Researchers had not limited their studies on X-ray images, they extend it to MRI. As an advantage, different types of brightness are shown for the same structures in an MRI scan [69]. The objective of this study [70] is to utilize DL techniques to deduce the malignancy of a bone tumor from magnetic resonance imaging (MRI) scans. The study's cohort consisted of 23 individuals, including 14 females and 9 males with ages ranging from 15 to 80 years. T1 and T2 weighted MRI scans are classified using two pretrained ResNet50 image classifiers. A clinical model is used to determine a tumor's likelihood of being malignant. The patient's clinical data and the results of the T1 and T2 classifiers are the model's inputs. It is a feed-forward neural network. Both classifiers achieved 95.00 percent accuracy throughout validation. The purpose of this scientific study [71] was to use the Turing test to evaluate an AI system's capacity to detect spine tumors. This paper suggests a fast R-CNN architecture for a DL-based tumor detection approach that, in two stages, first identifies the localization and the size of the bounding box around the lesion area and then creates region proposals using a region proposal network. The respondent's response was deemed incorrect if he failed to recognize the image that had been annotated by a person. The Turing test was regarded to have been passed if all mis-classification rates were >30 percent and the respondents could not tell the AI-detected tumor from the human-annotated one. As a result, The Turing test had an average misclassification rate of 51.2 percent.

### 3.7 Classification of pathological gait patterns

Human gait recognition and diagnosis become an active research area in AI [72, 73]. Ramirez et al. [74] suggest a sensor-based approaches for analyzing human gait. It introduces analytical methods in a framework for multivariate time series classification and interpretable anomaly detection for human gait study. Using a real-world clinical dataset in the field of biomechanical orthopedics to demonstrate the

application. Guo et al. [75] used a foot-pressure database gathered using the GAITRite walkway system to classify pathology-related variations in gait in young children. The classifier’s accuracy is improved by combining age information with the assessment of normal and abnormal gaits. With the help of this innovative approach, it may be possible to create a measure of pediatric gait abnormalities that is precise, affordable, and real-time. This measure would be able to inform clinicians about the effectiveness interventions and optimizing the giving treatment. With the use of a smart footbed equipped with multiple sensor arrays, Lee et al. [76] suggested a DL-based strategy for classifying different gait types. Gait data was collected by us using a smart insole that integrated a pressure sensor array, an acceleration sensor array, and a gyro sensor array. A deep convolution neural network was then used to retrieve gait pattern features (DCNN). The data acquired from each sensor array was then used to build an independent DCNN, which was used to extract a feature map. The feature maps were then integrated to create a fully linked network for classifying gait types. According to experimental results, the proposed approach demonstrated a remarkably high classification accuracy of over 90 percent for seven different gait patterns, including walking, fast walking, running, ascending stairs, descending stairs, climbing hills, and descending hills.

### 3.8 Prosthesis control

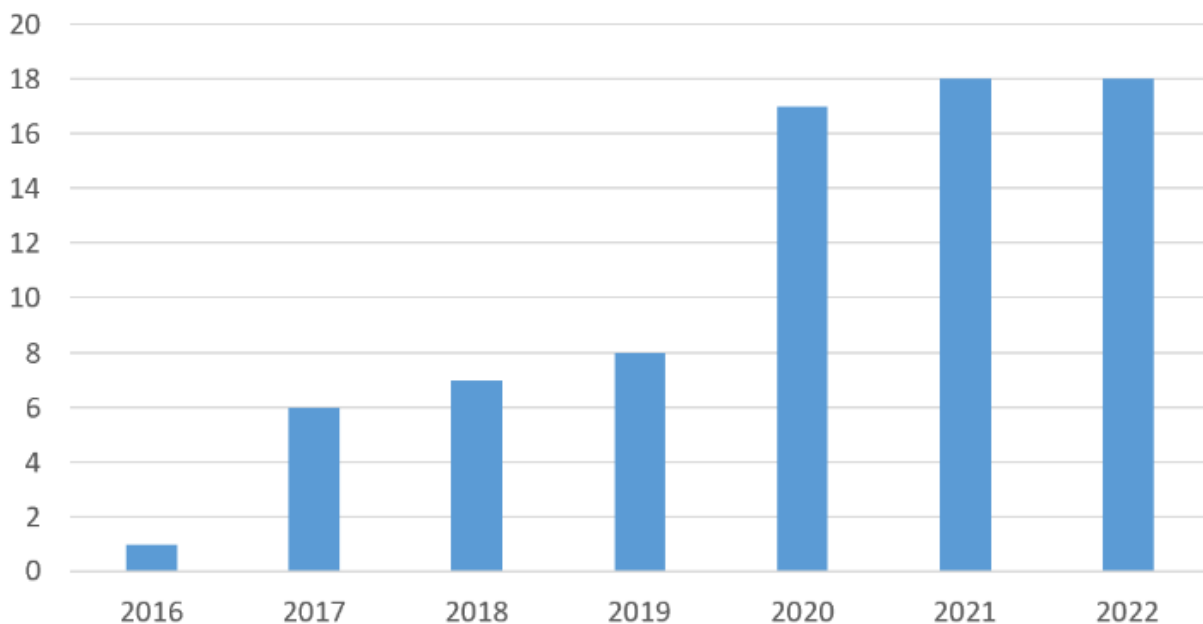
Researchers are exploring using AI and DL to regulate prosthetics as the field of smart prosthetics has advanced significantly in recent years. In the most of development studies, it was discovered that the CNN can be used to obtain faster and more efficient limb movements. The studies [60, 77] specialize and investigate their time to develop and validate a new approach based on ML techniques to analyse after a primary total hip arthroplasty. Age, race, gender, and comorbidity scores were the features having impact on model’s performance, with an AUROC of 0.87 and 0.71 for

LOS and payment respectively.

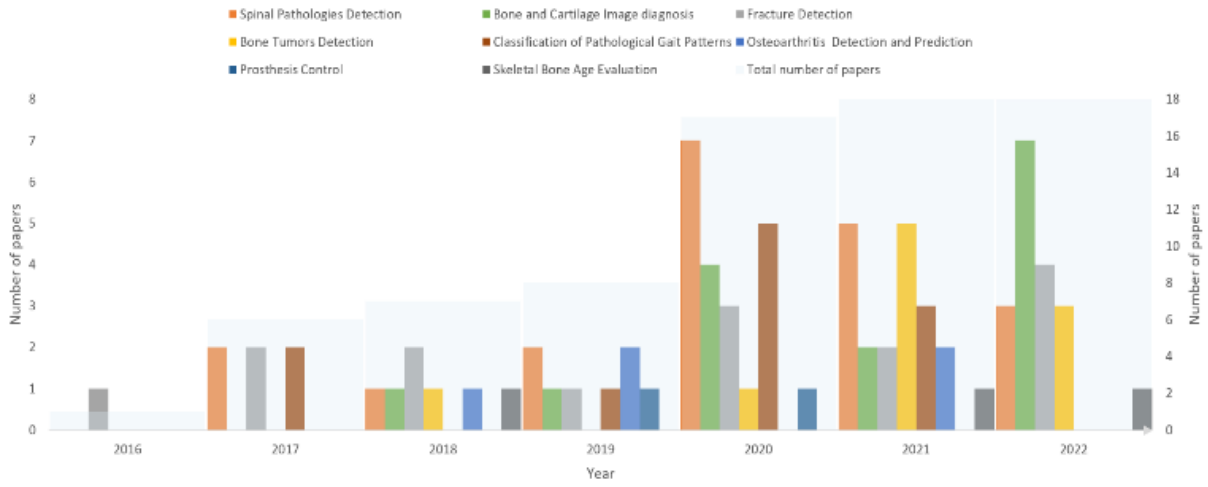
## 4. DISCUSSION

Recently, DL algorithms have shown remarkable accuracy in diagnosing orthopedic diseases from medical imaging. As a result, we conducted a comprehensive review of 75 studies that evaluated DL algorithms for identifying orthopedic abnormalities in medical images. As a result, the use of AI in medical imaging has evolved significantly over the past few years, offering new possibilities for improving medical diagnoses and treatments. GANs have shown promising results in medical imaging, as they can generate high-quality images with high resolution and low noise levels, which is critical for accurate diagnoses. Especially, GAN based architectures can help in developing personalized treatments for patients by generating images that reflect the unique characteristics of their conditions. GANs are will hopefully play an important part in medical imaging as their applications continue to evolve and expand.

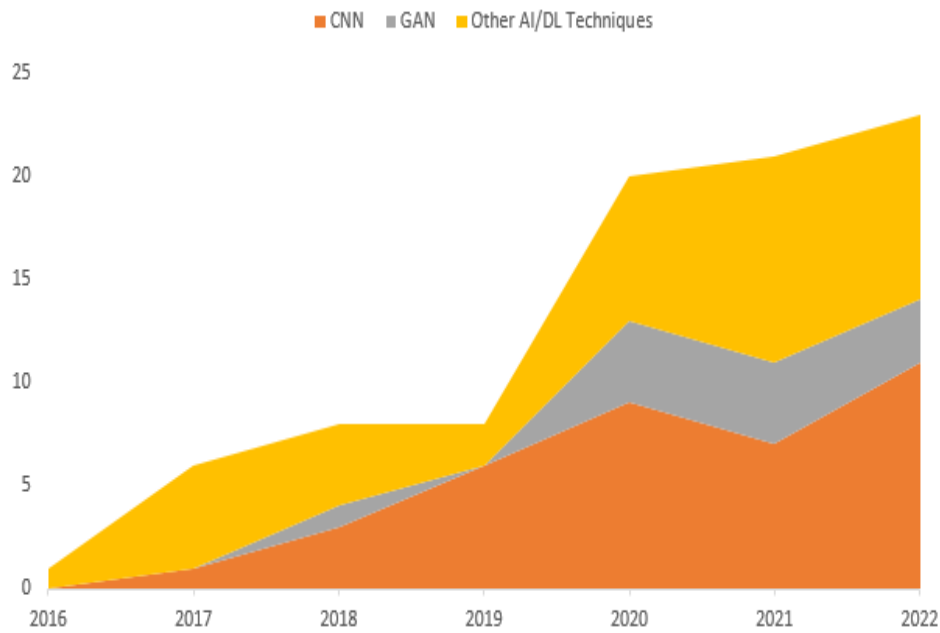
As shown in Figures 3-8, the field of orthopedics has seen significant advancements in recent years, and one of the most promising developments is the investigation of DL in the diagnosis of orthopedic anomalies. DL algorithms are capable of analyzing large scales of medical data and identifying patterns that are not immediately apparent to human experts. This can lead to more accurate and timely diagnoses, as well as more personalized treatment plans. By leveraging DL technology, orthopedic specialists can analyze X-rays, CT scans, and MRI images more efficiently, allowing for a more rapid and accurate diagnosis of conditions such as fractures, osteoporosis, and arthritis. The capacity of DL approaches to diagnosis, analyze and interpret complex medical data has the capacity to transform the orthopedics domain to a high level, enabling clinicians to provide more precise and effective care to patients with musculoskeletal disorders.



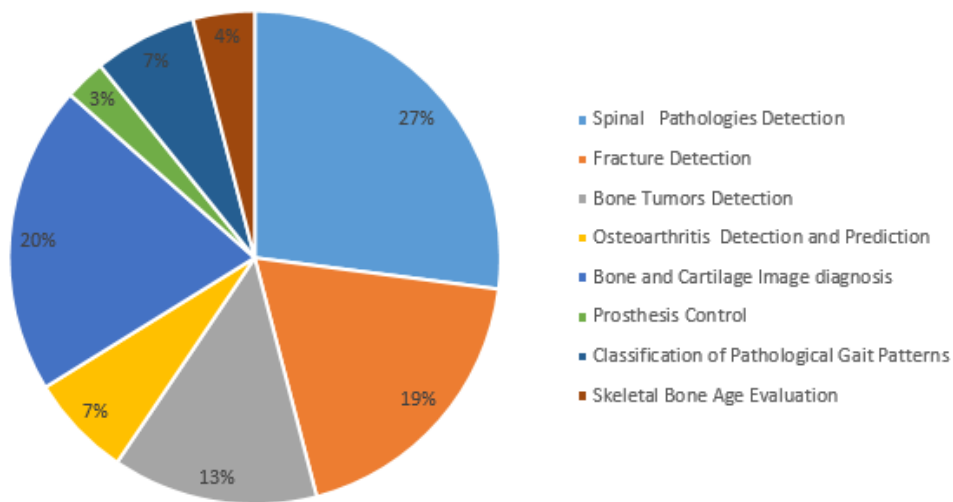
**Figure 3.** Year-on-year growth of AI publications in orthopedics



**Figure 4.** A temporal evolution chart provides a more in-depth view. The number of papers for each year is depicted and stratified based on the main classes of orthopedic application to display the trends.

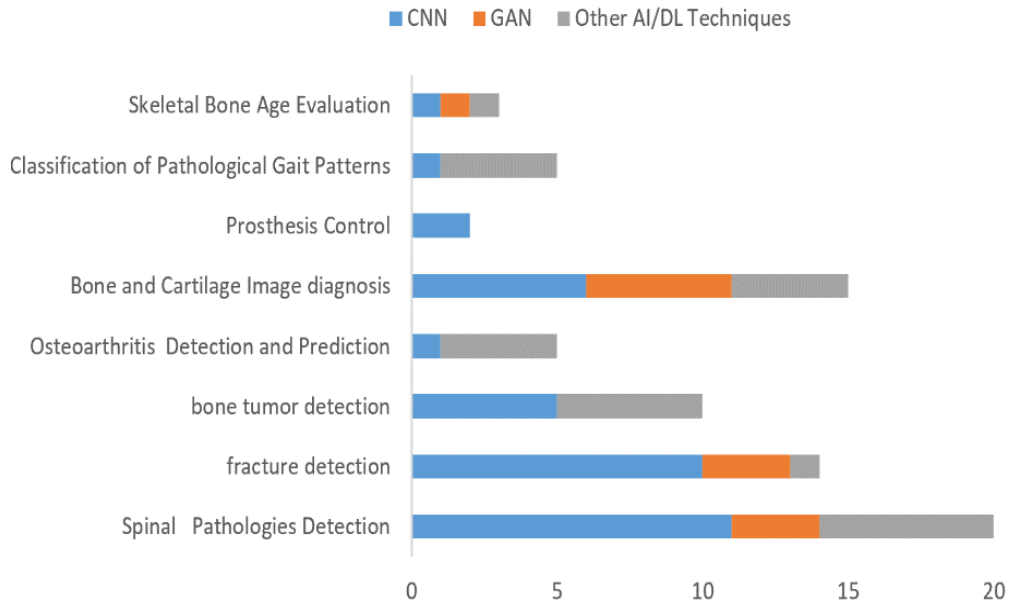


**Figure 5.** A temporal evolution chart illustrates the annual number of papers published since 2016 and included in the review due to their focus on machine learning applications in Orthopedics.

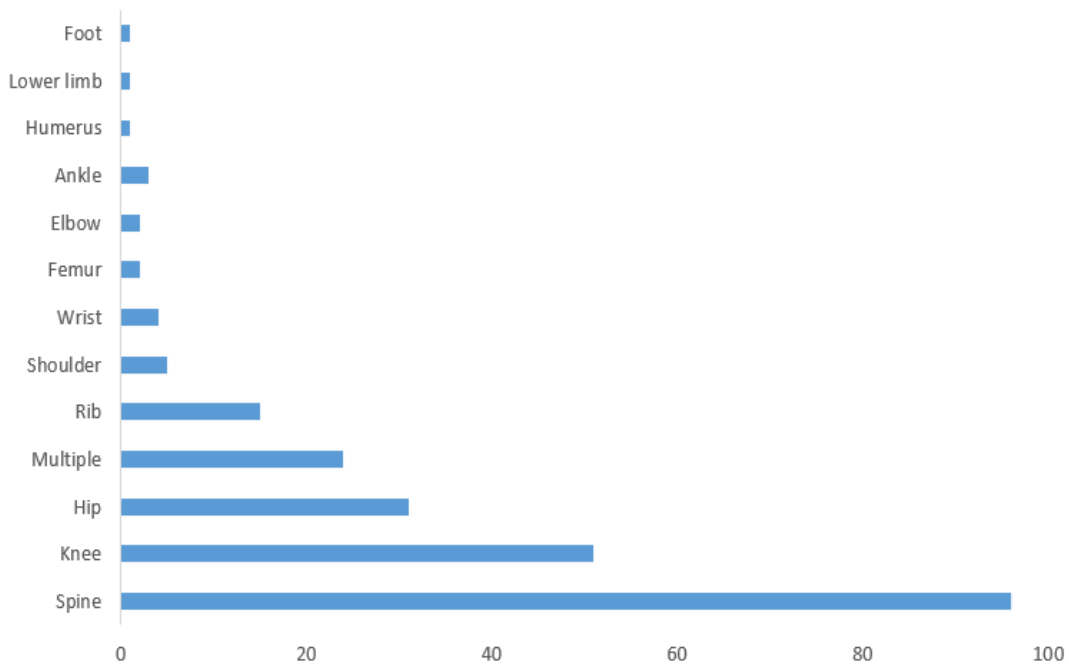


**Figure 6.** Percentage of the main diseases of orthopedic application according to reviewed studies in our work





**Figure 7.** Graph represents the number of papers included in this survey, categorized based on the machine learning techniques discussed and the main classes of orthopedic application.



**Figure 8.** A graph displays the number of papers published based on the area of the body.

Our review revealed substantial differences in study designs and outcome reporting. However, most studies utilized CNN and GAN-based algorithms, which have become the leading approach for AI in radiology. As shown in Table 2 we had listed a summary of the works performing GAN on their architectures. The studies that described overall performance showed moderate to outstanding results using DL strategies, although few studies compared their results with human or other AI techniques. Our objective was to provide a summary of the methodological aspects of the reviewed studies as there were previous reports have described various approaches in DL research for medical imaging, which has made it challenging to compare findings across different studies.

Regarding dataset, major studies used one modality of

medical images, others explore several modalities to have a computer aided system that train every image modality to give better diagnosis. Additionally, the studies we analyzed featured a broad spectrum of dataset sizes, ranging from 170 to 40,561 images. The methods used to divide these datasets into training, validation, and testing sets were also diverse, varying from cross-validation to using an independent test set. This diversity in methodology presents challenges when trying to compare results across articles, because one model may exhibit different performance. relative to the size of the dataset, according to the nature of the images, percentage of anomalous data and of course based on cross-validation measured to a true external test set.

**Table 2.** Summary of the works performing GAN

[Ref](Year)	Aim of the study	Data type	Approach	Results
[11] (2021)	Spinal Pathology Detection	MRI	Propose one-step approaches based on the GAN architecture which are cable and powerful to segment the components of the spine namely: discs, vertebrae.	Recall = 0.857, Precision = 0.888
[19] (2022)	Spinal Pathology Detection	X-ray	The proposed method employs a GAN in a semi-supervised manner to train on mild to severe scoliosis cases. The GAN serves as an upstream task to learn scoliosis representations and a downstream task to accurately classify between normal and scoliotic states.	Negative predictive value = 0.950
[20] (2018)	Spinal Pathology Detection	MRI	Introducing a model SpineGAN is characterized by its architecture structured by components and AE modules that allows it to solve the problem of the great variety and variability of complex spinal structures.	Accuracy of 96.2 percent
[21] (2020)	Spinal Pathology Detection	CT, MR images	Propose a new method that synthesizes CT images of the lumbar spine using a fully unsupervised approach. The method utilizes a T2-weighted MRI that is acquired for diagnostic purposes to generate images that can be used in image-guided surgical procedures.	Dice score of $83 \pm 1.6$
[38] (2021)	Fracture Detection	X-ray	A proposed unsupervised anomaly detection method is the Res-UNetGAN Network, which utilizes a Generative Adversarial Network. This network combines a ResNet50 and UNet architecture to form an autoencoder structure that is capable of learning features from the data. The generative network is used to detect anomalies in an unsupervised manner.	Res-UNetGAN: 0.92 GANomaly: 0.81 Skip-GANomaly: 0.90 CVAE-GAN-Based: 0.86 EGBAD: 0.80
[39] (2020)	Fracture Detection	X-ray	Comparative study between GAN and AE models on anomaly detection.	DCGAN: 0.53 BiGAN: 0.54 AlphaGAN: 0.60 VAE: 0.48 CAE: 0.57
[55] (2022)	Bone and Cartilage Image diagnosis	Microscopic images	One potential technique to enhance the precision of cell classification in bone marrow aspirate smears is to create synthetic data using a three-stage architecture grounded on the GAN approach. The generated synthetic data can be integrated with the original data to support in elevating the precision of cell classification.	Accuracy = 96 percent
[57] (2020)	Bone and Cartilage Image Diagnosis	Ultrasound Images	A GAN model has been suggested, which incorporates two convolutional blocks known as self-projection and self-attention blocks. These blocks are utilized to generate authentic B-mode bone ultrasound images and segmented bone masks.	Acc = 85 percent
[58] (2020)	Bone and Cartilage Image Diagnosis	ultrasound images	A proposition has been made to introduce a computational technique, utilizing GAN architecture, that enables the swift and real-time segmentation of bone shadow images. This can subsequently be integrated into a multi-feature guided CNN architecture, allowing for precise and real-time bone surface segmentation.	Mean dice coefficient of 93 percent
[61] (2021)	Skeletal Bone Age Evaluation	Photogrammetric scans	A proposed method to enhance BAA training using DL is to incorporate a combination of CNN, GAN, and One-shot Temporal Dependency (OTD) into both steps of the pretraining and training architecture.	Mean Average Error of 4.23

AI models showed better performance with high-quality, large-scale datasets. In our review, it was acknowledged that the size of the dataset plays a significant role. The validity of studies that use small datasets may come under scrutiny. Nevertheless, research utilizing the largest available datasets remains a valuable asset for future investigations.

The non-use of a true external test set degrades the value of the article. External validation of algorithms is primordial due

to the fact that the effectiveness of models may differ when applied to data from different hospitals or environments. The lack of external validation in many studies could significantly impact the trustworthiness of AI systems in real-world clinical environments. External validation, which involves testing the performance of AI systems on independent datasets or in different clinical settings, is crucial for assessing their generalizability and reliability. Without adequate validation,

there is a risk of overestimating the performance of AI systems based on their performance in limited experimental conditions, leading to inflated expectations and potential failures when deployed in real-world clinical settings. Furthermore, without external validation, there is a lack of evidence to support the effectiveness and safety of AI systems across diverse patient populations and clinical scenarios. This could erode trust among healthcare professionals and patients, hindering the adoption and acceptance of AI systems in clinical practice. Therefore, ensuring robust external validation is essential for building confidence in the reliability and effectiveness of AI systems in real-world healthcare settings.

Based on this criteria, we surprisingly found that only few studies explored an external test set, which could be associated to data availability challenge. Data availability presents significant challenges for developing and testing AI models in healthcare, particularly in niche areas like orthopedics. Limited access to high-quality, annotated datasets can hinder model training and validation, leading to potential biases and suboptimal performance. To improve data sharing within the research community, initiatives promoting open data repositories and standardized sharing protocols could facilitate access to diverse datasets while addressing privacy concerns. Incentivizing data sharing through funding and publication policies could also encourage researchers to contribute their datasets. Overall, fostering a culture of collaboration and transparency in data sharing is crucial for advancing AI-driven healthcare and improving patient outcomes.

Although the heterogeneous study design we could report the main tasks treated by most of studies:

- Many studies focused on using imaging to make a diagnosis.
- The spine being the most commonly researched musculoskeletal area.
- Fracture and cartilage diseases diagnosis are an emerging area of interest.

GAN-based models were explored, in several reported studies included in this review, interpreting imaging results to make a diagnosis is a popular use of AI, due to the abundance of structured data received during imaging and the ease of creating GAN models to analyze it. Radiology has seen a significant rise in the usage of AI for interpreting scans, particularly in fracture detection and spine pathology diagnosis. The majority of studies have focused on the spine, possibly due to the collaboration between radiology and neurosurgery in managing spinal issues. However, more attention should be given to other subspecialty areas such as the hand, hip, and knee, which also benefit from AI intervention.

A remarkable amount of research investigated through the literature search associated to the applicability of CNNs and GANs as their primary DL technique. They compared the performance of the implemented models, and they considerable observed increase in performance with the frequent methods of pretraining and data augmentation. This large variation in performance (AUC 40 to 94 percent) when using diverse CNN or GAN architecture trained may be due to the integration of some specific component to the principle architecture like AE. The illustration and verification of crucial image regions in decision-making can also be achieved through the utilization of either computational attention mechanisms (CAMs) or salience maps.

The standard form of frequent performance measures are AUC, accuracy, sensitivity, and specificity, most of studies

included in our review some measures is lacking and only a few numbers of studies had reported performance solely based on the AURC.

Hence, reporting results put us in a difficulty in analyzing parameter metrics, interpreting and comparing performance with other DL models or doctors.

DL models have emerged as powerful tools in orthopedic diagnostics, offering distinct advantages over traditional methods in terms of accuracy, efficiency, and cost-effectiveness. In terms of accuracy, DL models, particularly CNNs, have demonstrated remarkable proficiency in analyzing medical images with a level of precision that rivals or even surpasses that of human experts. These models excel in detecting subtle abnormalities and patterns in imaging data, leading to more accurate diagnoses and treatment recommendations. Furthermore, DL models are highly efficient, capable of automating image analysis tasks that would traditionally require significant manual effort and time from skilled radiologists or orthopedic specialists. By swiftly processing large volumes of medical imaging data, these models expedite the diagnostic process, allowing for quicker turnaround times and reducing the burden on healthcare professionals. Moreover, DL models can offer cost-effective solutions in orthopedic diagnostics. While the initial investment in developing and implementing these models may be substantial, once deployed, they can significantly reduce long-term costs associated with manual image interpretation, such as labor expenses and potential errors. Additionally, by streamlining workflows and improving diagnostic accuracy, DL models can contribute to better resource allocation and more efficient use of healthcare resources. Overall, the comparison between DL models and traditional methods underscores the transformative potential of AI in orthopedic diagnostics. These models not only offer superior accuracy and efficiency but also present cost-effective solutions that can enhance patient care and optimize healthcare delivery in orthopedic practice.

Generally, the studies reviewed in our paper had medium to excellent performance in diagnosing orthopedic anomalies from different image modalities, with the vast majority using CNN and GAN-based architecture. This conducts that AI models may have the ability to have a real clinical role as a doctor assistance in the near future in diagnosing pathologies using imaging.

In this study, few papers had performed comparable models to some radiologists, highlighting that AI may have a decrease in benefits for more experienced physicians. Hence, AI algorithms could be reasonably integrated usefully in the diagnosis process. The various studies included in this review have demonstrated some level of uncertainty, but the utilization of DL algorithms can offer multiple options and corresponding levels of confidence to assist a radiologist in making their final diagnosis. Some of the studies in the review also use CAMs and saliency maps, which provide insight into the decision-making process by highlighting crucial areas of the image that can enhance the results and facilitate human decision-making.

In this review, most reviewed studies were concentrated on diagnosis orthopedic anomalies through which AI systems specially DL models CNN and GAN-based architectures. Medical image modalities were the main input source in the studies. Indeed, exploring CAD systems have the capacity to diagnose some diseases notably cartilage pathology, bone deformation, tumors identification based on algorithms with a

high accuracy. Other studies displayed CAD systems has the ability to identify spinal pathology by implementing CNN and GAN models capable of segmenting discs and vertebrae. Nonetheless, they technically-sound have been shown, and the results are noteworthy when specifically looking at spinal diagnosis. Additionally, some studies have demonstrated a notably high level of accuracy in image diagnosis. by having radiologists note for every input.

Collectively, studies have reported accuracy rates ranging from 40 to 94 percent using both conventional and generative techniques in the diagnosis and treatment of bones. These techniques have the potential to streamline time-consuming tasks and provide new insights from previously unused data. The most extensively researched task with the best performance outcomes has been the identification and grading of organs. Nevertheless, there have been successful efforts to incorporate multiple medical imaging modalities into CAD systems, yielding positive results.

The study has several limitations. Firstly, the diverse methodologies, data sources, and outcomes among the studies made it impossible to conduct a meta-analysis. Secondly, the search was limited to English manuscripts only, meaning that articles written in other languages that fit the inclusion criteria may have been missed.

## 5. CONCLUSIONS

Artificial Intelligence is significantly influencing medical research and the care of patients, with numerous applications in the diagnosis of orthopedic conditions. This study systematically reviewed 75 research studies that utilized AI and DL techniques, with a specific emphasis on the use of CNN and GAN for orthopedic diagnosis.

The results indicated that the AI algorithms showed promising results, with favorable outcomes. However, there was a significant variation in study designs, which made it challenging to accurately evaluate and compare the performance of different models.

To further enhance the understanding of AI's performance, future studies should aim to adopt consistent training and testing methods and provide more comprehensive and transparent reporting of their methods and outcomes. Ultimately, ensuring clinical relevance in the development of AI models is paramount for their effective integration into real-world healthcare settings. Future studies should focus on testing models in ways that align with clinical workflows and patient care outcomes. Collaboration between AI developers, healthcare providers, and regulators is key to defining relevant performance metrics and standards. By prioritizing clinical relevance, researchers can ensure that AI models have a meaningful impact on patient care and clinical decision-making.

The implementation of AI in orthopedic diagnosis necessitates thorough cost-benefit analyses to justify the investment in this technology. While AI has the potential to enhance diagnostic accuracy, improve patient outcomes, and optimize resource allocation, its adoption requires significant financial investment, infrastructure development, and training of personnel. Conducting cost-benefit analyses can provide insights into the potential return on investment (ROI) of AI implementations, taking into account factors such as reduced healthcare costs, improved efficiency, and enhanced patient satisfaction. Additionally, assessing the long-term financial

implications, including maintenance costs and scalability, is essential for ensuring the sustainability of AI implementations. By quantifying the expected benefits against the associated costs, healthcare stakeholders can make informed decisions regarding the adoption of AI in orthopedic diagnosis, ultimately maximizing value for both patients and healthcare systems.

## ACKNOWLEDGMENT

We want to convey my heartfelt appreciation to my supervisors and colleagues for their invaluable contributions and support, which have significantly enhanced the content and direction of this article. Their guidance and insights played a crucial role in shaping its development.

## REFERENCES

- [1] Xie, Z. (2016). Machine learning for efficient recognition of anatomical structures and abnormalities in biomedical images (Doctoral dissertation, Imperial College London).
- [2] Haq, R., Schmid, J., Borgie, R., Cates, J. (2020). Deformable multisurface segmentation of the spine for orthopedic surgery planning and simulation. *Journal of Medical Imaging*, 7(1): 015002. <https://doi.org/10.1117/1.JMI.7.1.015002>
- [3] Vergari, C., Skalli, W., Gajny, L. (2020). A convolutional neural network to detect scoliosis treatment in radiographs. *International Journal of Computer Assisted Radiology and Surgery*, 15(6): 1069-1074. <https://doi.org/10.1007/s11548-020-02173-4>
- [4] Fraiwan, M., Audat, Z., Fraiwan, L., Manasreh, T. (2022). Using deep transfer learning to detect scoliosis and spondylolisthesis from X-ray images. *Plos One*, 17(5): e0267851. <https://doi.org/10.1371/journal.pone.0267851>
- [5] Zhang, L., Zhang, J., Gao, S. (2022). Region-based convolutional neural network-based spine model positioning of X-ray images. *BioMed Research International*, 2022: 7512445. <https://doi.org/10.1155/2022/7512445>
- [6] Yang, J., Zhang, K., Fan, H., Huang, Z., Xiang, Y., Yang, J., He, L., Zhang, L., Yang, Y., Li, R. (2019). Development and validation of deep learning algorithms for scoliosis screening using back images. *Communications Biology*, 2(1): 1-8. <https://doi.org/10.1038/s42003-019-0635-8>
- [7] Galbusera, F., Niemeyer, F., Wilke, H.J., Bassani, T., Casaroli, G., Anania, G., Costa, F., Brayda-Bruno, M., Sconfienza, L.M. (2019). Fully automated radiological analysis of spinal disorders and deformities: A deep learning approach. *European Spine Journal*, 28(5): 951-960. <https://doi.org/10.1007/s00586-019-05944-z>
- [8] Gao, F., Liu, S., Zhang, X., Wang, X., Zhang, J. (2021). Automated grading of lumbar disc degeneration using a push-pull regularization network based on MRI. *Journal of Magnetic Resonance Imaging*, 53(3): 799-806. <https://doi.org/10.1002/jmri.27400>
- [9] Lewandrowski, K.U., Muraleedharan, N., Eddy, S.A., Sobti, V., Reece, B.D., ramírez León, J.F., Shah, S. (2020). Feasibility of deep learning algorithms for reporting in routine spine magnetic resonance imaging. *International Journal of Spine Surgery*, 14(s3): S86-S97.

- <https://doi.org/10.14444/7131>
- [10] Hallinan, J.T.P.D., Zhu, L., Yang, K., Makmur, A.R.A., Algazwi, D.A.R., Thian, Y.L., Lau, S., Choo, Y.S., Eide, S.E., Yap, Q.V. (2021). Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine MRI. *Radiology*, 300(1): 130-138. <https://doi.org/10.1148/radiol.2021204289>
- [11] He, J., Liu, W., Wang, Y., Ma, X., Hua, X.S. (2021). Spineone: A one-stage detection framework for degenerative discs and vertebrae. In 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, pp. 1331-1334. <https://doi.org/10.48550/arXiv.2110.15082>
- [12] Du, X., Lin, T.Y., Jin, P., Ghiasi, G., Tan, M., Cui, Y., Le, Q.V., Song, X. (2020). Spinenet: Learning Scale-Permuted backbone for recognition and localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA. pp. 11592-11601. <https://doi.org/10.1109/CVPR42600.2020.01161>
- [13] Wu, H., Bailey, C., Rasoulinejad, P., Li, S. (2017). Automatic landmark estimation for adolescent idiopathic scoliosis assessment using boostnet. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, pp. 127-135. [https://doi.org/10.1007/978-3-319-66182-7\\_15](https://doi.org/10.1007/978-3-319-66182-7_15)
- [14] Yeh, Y.C., Weng, C.H., Huang, Y.J., Fu, C.J., Tsai, T.T., Yeh, C.Y. (2021). Deep learning approach for automatic landmark detection and alignment analysis in whole spine lateral radiographs. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-87141-x>
- [15] Zhang, J., Li, H., Lv, L., Zhang, Y. (2017). Computer-aided cobb measurement based on automatic detection of vertebral slopes using deep neural network. *International Journal of Biomedical Imaging*, 2017: 9083916. <https://doi.org/10.1155/2017/9083916>
- [16] Cina, A., Bassani, T., Panico, M., Luca, A., Masharawi, Y., Brayda-Bruno, M. (2021). 2-step deep learning model for landmarks localization in spine radiographs. *Scientific Reports*, 11(1): 1-12. <https://doi.org/10.1038/s41598-021-89102-w>
- [17] Yi, J., Wu, P., Huang, Q., Qu, H. (2020). Vertebra focused landmark detection for scoliosis assessment. In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, pp. 736-740. <https://doi.org/10.1109/ISBI45749.2020.9098675>
- [18] Caesarendra, W., Rahmaniar, W., Mathew, J., Thien, A. (2022). AutoSpine-Net: Spine detection using convolutional neural networks for cobb angle classification in adolescent idiopathic scoliosis. In Proceedings of the 2nd International Conference on Electronics, Biomedical Engineering, and Health Informatics, pp. 547-556. [https://doi.org/10.1007/978-981-19-1804-9\\_41](https://doi.org/10.1007/978-981-19-1804-9_41)
- [19] Lee, W., Shin, K., Lee, J., Yoo, S.J., Yoon, M.A., Choi, Y.W., Hong, G.S., Kim, N., Paik, S. (2022). Diagnosis of scoliosis using chest radiographs with a semi-supervised generative adversarial network. *Journal of the Korean Society of Radiology*, 83(6): 1298-1311. <https://doi.org/10.3348/jksr.2021.0146>
- [20] Han, Z., Wei, B., Mercado, A., Leung, S., Li, S. (2018). Spine-GAN: Semantic segmentation of multiple spinal structures. *Medical Image Analysis*, 50: 23-35. <https://doi.org/10.1016/j.media.2018.08.005>
- [21] Oulbacha, R., Kadoury, S. (2020). MRI to CT synthesis of the lumbar spine from a Pseudo-3D cycle GAN. In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, pp. 1784-1787. <https://doi.org/10.1109/ISBI45749.2020.9098421>
- [22] Moon, G., Kim, S., Kim, W., Kim, Y., Jeong, Y., Choi, H.S. (2022). Computer Aided Facial Bone Fracture Diagnosis (CA-FBFD) system based on object detection model. *IEEE Access*, 10: 79061-79070. <https://doi.org/10.1109/ACCESS.2022.3192389>
- [23] Sathish Kumar, L., Prabu, A., Pandimurugan, V., Rajasoundaran, S., Malla, P.P., Routray, S. (2022). A comparative experimental analysis and deep evaluation practices on human bone fracture detection using x ray images, *Concurrency and Computation: Practice and Experience*, 34(26). <https://doi.org/10.1002/cpe.7307>
- [24] Burns, J.E., Yao, J., Summers, R.M. (2017). Vertebral body compression fractures and bone density: Automated detection and classification on CT images. *Radiology* 284(3): 788. <https://doi.org/10.1148/radiol.2017162100>
- [25] Kruse, C., Eiken, P., Vestergaard, P. (2017). Machine learning principles can improve hip fracture prediction. *Calcified Tissue International*, 100(4): 348-360. <https://doi.org/10.1007/s00223-017-0238-7>
- [26] Mehr, G. (2020). Automating abnormality detection in musculoskeletal radiographs through deep learning. <https://doi.org/10.48550/arXiv.2010.12030>
- [27] Ounasser, N., Rhanoui, M., Mikram, M., El Asri, B. (2023). Anomaly detection in orthopedic musculoskeletal radiographs using deep learning. In International Congress on Information and Communication Technology, London, pp. 93-102. [https://doi.org/10.1007/978-981-99-3243-6\\_8](https://doi.org/10.1007/978-981-99-3243-6_8)
- [28] Ounasser, N., Rhanoui, M., Mikram, M., El Asri, B., Sekkaki, A. (2023). Enhancing computer-assisted bone fractures diagnosis in musculoskeletal radiographs based on generative adversarial networks. *International Journal of Advanced Computer Science and Applications*, 14(7). <https://doi.org/10.14569/IJACSA.2023.01407104>
- [29] Chung, S.W., et al. (2018). Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthopaedica*, 89(4): 468-473. <https://doi.org/10.1080/17453674.2018.1453714>
- [30] Cheng, L.W., Chou, H.H., Huang, K.Y., Hsieh, C.C., Chu, P.L., Hsieh, S.Y. (2022). Automated diagnosis of vertebral fractures using radiographs and machine learning. In International Conference on Intelligent Computing, Xi'an, China, pp. 726-738. [https://doi.org/10.1007/978-3-031-13870-6\\_59](https://doi.org/10.1007/978-3-031-13870-6_59)
- [31] Goyal, M., Malik, R., Kumar, D., Rathore, S. (2020). Musculoskeletal abnormality detection in medical imaging using GnCNNr (Group Normalized Convolutional Neural Networks with Regularization). *SN Computer Science*, 1(6). <https://doi.org/10.1007/s42979-020-00340-7>
- [32] Ounasser, N., Rhanoui, M., Mikram, M., El Asri, B. (2024). Advancing medical imaging with GAN-based anomaly detection. *Indonesian Journal of Electrical Engineering and Computer Science*, 35. <http://doi.org/10.11591/ijeecs.v35.i1.pp570-582>
- [33] Ounasser, N., Rhanoui, M., Mikram, M., El Asri, B.

- (2021). Generative and autoencoder models for large-scale multivariate unsupervised anomaly detection. In *Networking, Intelligent Systems and Security: Proceedings of NISS 2021*, pp. 45-58. [https://doi.org/10.1007/978-981-16-3637-0\\_4](https://doi.org/10.1007/978-981-16-3637-0_4)
- [34] Regnard, N.E., Lanseur, B., Ventre, J., Ducarouge, A., Clovis, L., Lassalle, L., Lacave, E., Grandjean, A., Lambert, A., Dallaudière, B., Berge, J., Troussset, Y. (2022). Assessment of performances of a deep learning algorithm for the detection of limbs and pelvic fractures, dislocations, focal bone lesions, and elbow effusions on trauma X-rays. *European Journal of Radiology*, 154: 110447. <https://doi.org/10.1016/j.ejrad.2022.110447>
- [35] Abreu Dias, D.D. (2019). Musculoskeletal abnormality detection on x-ray using transfer learning (2019). [https://repositori.upf.edu/bitstream/handle/10230/42540/De\\_Abreu\\_2019.pdf?sequence=1](https://repositori.upf.edu/bitstream/handle/10230/42540/De_Abreu_2019.pdf?sequence=1).
- [36] Kim, D., MacKinnon, T. (2018). Artificial intelligence in fracture detection: Transfer learning from deep convolutional neural networks. *Clinical Radiology*, 73(5): 439-445. <https://doi.org/10.1016/j.crad.2017.11.015>
- [37] Spahr, A., Bozorgtabar, B., Thiran, J.P. (2021). Self-taught semi-supervised anomaly detection on upper limb X-rays. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 1632-1636. <https://doi.org/10.48550/arXiv.2102.09895>
- [38] Song, S., Yang, K., Wang, A., Zhang, S., Xia, M. (2021). A mura detection model based on unsupervised adversarial learning. *IEEE Access*, 9: 49920-49928. <https://doi.org/10.1109/ACCESS.2021.3069466>
- [39] Davletshina, D., Melnychuk, V., Tran, V., Singla, H., Berrendorf, M., Faerman, E., Fromm, M., Schubert, M. (2020). Unsupervised anomaly detection for X-ray images. *arXiv preprint arXiv:2001.10883*. <https://doi.org/10.48550/arXiv.2001.10883>
- [40] Tiulpin, A., Thevenot, J., Rahtu, E., Lehenkari, P., Saarakkala, S. (2018). Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-Based Approach. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-20132-7>
- [41] Jamshidi, A., Pelletier, J.P., Martel-Pelletier, J. (2019). Machine-learning-based patient-specific prediction models for knee osteoarthritis. *Nature Reviews Rheumatology*, 15(1): 49-60. <https://doi.org/10.1038/s41584-018-0130-5>
- [42] Jakaite, L., Schetinina, V., Hladvka, J., Minaev, S., Ambia, A., Krzanowski, W. (2021). Deep learning for early detection of pathological changes in X-ray bone microstructures: Case of osteoarthritis. *Scientific Reports*, 11: 2294 (2021). <https://doi.org/10.1038/s41598-021-81786-4>
- [43] Wang, Y., Wang, X., Gao, T., Du, L. (2021). An automatic knee osteoarthritis diagnosis method based on deep learning: Data from the osteoarthritis initiative. *Journal of Healthcare Engineering*, 2021(1). <https://doi.org/10.1155/2021/5586529>
- [44] Lim, J., Kim, J., Cheon, S. (2019). A deep neural network-based method for early detection of osteoarthritis using statistical data. *International Journal of Environmental Research and Public Health*, 16(7): 1281. <https://doi.org/10.3390/ijerph16071281>
- [45] Alotaibi, G., Awawdeh, M., Farook, F.F., Aljohani, M., Aldhafiri, R.M., Aldhoayan, M. (2022). Artificial intelligence (AI) diagnostic tools: Utilizing a convolutional neural network (CNN) to assess periodontal bone level radiographically—A retrospective study. *BMC Oral Health*, 22(1). <https://doi.org/10.1186/s12903-022-02436-3>
- [46] Bien, N., Rajpurkar, P., Ball, R.L., Irvin, J., Park, A., Jones, E., Bereket, M., Patel, B.N., Yeom, K.W., Shpanskaya, K., Halabi, T., Zucker, J.D., Riley, C., Pankanti, A., Domínguez, A., Beers, A., Rodriguez, R. G., Taylor, B., Mott, J., Wong, C., Hecht, M.L., Elmore, G., Sandler, R.S., Friedman, D., Aradi, A.R., Dreyer, K.J. (2018). Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. *PLoS Medicine*, 15(11): e1002699. <https://doi.org/10.1371/journal.pmed.1002699>
- [47] Gatti, A.A., Maly, M.R. (2021). Automatic knee cartilage and bone segmentation using multi-stage convolutional neural networks: Data from the osteoarthritis initiative. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 34(6): 859-875. <https://doi.org/10.1007/s10334-021-00934-z>
- [48] Brui, E., Efimtcev, A.Y., Fokin, V.A., Fernandez, R., Levchuk, A.G., Ogier, A.C., Samsonov, A.A., Mattei, J.P., Melchakova, I.V., Bendahan, D. (2020). Deep learning based fully automatic segmentation of wrist cartilage in MR images. *NMR in Biomedicine*, 33(8): e4320. <https://doi.org/10.1002/nbm.4320>
- [49] Yao, J., Guo, Z., Yu, W., Wang, G. (2022). Enhanced deep residual network for bone classification and abnormality detection. *Medical Physics*, 49(11). <https://doi.org/10.1002/mp.15966>
- [50] Basavaraja, P.H., Ganesarathinam, S. (2022). An ensemble-of deep learning model with optimally selected features for osteoporosis detection from bone X-ray Images. *International Journal of Intelligent Engineering & Systems*, pp. 194-206. <https://doi.org/10.22266/ijies2022.1031.18>
- [51] Singh, G., Anand, D., Cho, W., Joshi, G.P., Son, K.C. (2022). Hybrid deep learning approach for automatic detection in musculoskeletal radiographs. *Biology*, 11(5): 665. <https://doi.org/10.3390/biology11050665>
- [52] He, M., Wang, X., Zhao, Y. (2021). A calibrated deep learning ensemble for abnormality detection in musculoskeletal radiographs. *Scientific Reports*, 11(1): 1-11. <https://doi.org/10.1038/s41598-021-88578-w>
- [53] Noguchi, S., Nishio, M., Yakami, M., Nakagomi, K., Togashi, K. (2020). Bone segmentation on whole-body CT using convolutional neural network with novel data augmentation techniques. *Computers in Biology and Medicine*, 121: 103767. <https://doi.org/10.1016/j.compbiomed.2020.103767>
- [54] Varma, M., Lu, M., Gardner, R., Dunnmon, J., Khandwala, N., Rajpurkar, P., Long, J., Beaulieu, C., Shpanskaya, K., Fei-Fei, L., Botkin, J.R. (2019). Automated abnormality detection in lower extremity radiographs using deep learning. *Nature Machine Intelligence*, 1(12): 578-583. <https://doi.org/10.1038/s42256-019-0126-0>
- [55] Hazra, D., Byun, Y.C., Kim, W.J. (2022). Enhancing classification of cells procured from bone marrow aspirate smears using generative adversarial networks and sequential convolutional neural network. *Computer Methods and Programs in Biomedicine*, 224: 107019. <https://doi.org/10.1016/j.cmpb.2022.107019>

- [56] Park, C., Kang, J.W., Lee, D.E., Son, W., Lee, S.M., Park, C. (2022). Deep learning approaches for bone marrow edema detection and interpretation in dual energy CT. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4189440>
- [57] Alsinan, A.Z., Patel, V.M., Hacihaliloglu, I. (2020). Bone shadow segmentation from ultrasound data for orthopedic surgery using GAN. *International Journal of Computer Assisted Radiology and Surgery*, 15(9): 1477-1485. <https://doi.org/10.1007/s11548-020-02221-z>
- [58] Alsinan, A.Z., Rule, C., Vives, M., Patel, V.M., Hacihaliloglu, I. (2020). Gan-based realistic bone ultrasound image and label synthesis for improved segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Lima, Peru, pp. 795-804. [https://doi.org/10.1007/978-3-030-59725-2\\_77](https://doi.org/10.1007/978-3-030-59725-2_77)
- [59] Zhou, Y., Rakkunedeth, A., Keen, C., Knight, J. (2022). Wrist ultrasound segmentation by deep learning. In *International Conference on Artificial Intelligence in Medicine*, Halifax, NS, Canada, pp. 230-237. [https://doi.org/10.1007/978-3-031-09342-5\\_22](https://doi.org/10.1007/978-3-031-09342-5_22)
- [60] Ren, X., Li, T., Yang, X., Wang, S., Ahmad, S., Xiang, L., Stone, S.R., Li, L., Zhan, Y., Shen, D. (2018). Regression convolutional neural network for automated pediatric bone age assessment from hand radiograph. *IEEE Journal of Biomedical and Health Informatics*, 23(5): 2030-2038. <https://doi.org/10.1109/JBHI.2018.2876916>
- [61] Su, L., Fu, X., Hu, Q.M. (2021). Generative adversarial network based data augmentation and gender-last training strategy with application to bone age assessment. *Computer Methods and Programs in Biomedicine*, 212: 106456. <https://doi.org/10.1016/j.cmpb.2021.106456>
- [62] Li, S., Liu, B., Li, S., Zhu, X., Yan, Y., Zhang, D., Zheng, Y. (2022). A deep learning-based computer-aided diagnosis method of X-ray images for bone age assessment. *Complex & Intelligent Systems*, 8(3): 1929-1939. <https://doi.org/10.1007/s40747-021-00376-z>
- [63] Shrivastava, D., Sanyal, S., Maji, A.K. (2020). Bone Cancer detection using machine learning techniques. In *Smart Healthcare for Disease Diagnosis and Prevention*, Pathum Thani, Thailand, pp: 175-183. <https://doi.org/10.1109/ICCMSO58359.2022.00068>
- [64] B. Vandana, Alva, S.R. (2021). Deep learning based Automated tool for cancer diagnosis from bone histopathology images. In *2021 International Conference on Intelligent Technologies (CONIT)*, Hubli, India, pp. 1-8. <https://doi.org/10.1109/CONIT51480.2021.9498367>
- [65] Sharma, A., Yadav, D.P., Garg, H., Kumar, M., Sharma, B. (2021). Bone cancer detection using feature extraction based machine learning model. *Computational and Mathematical Methods in Medicine*, 2021. <https://doi.org/10.1155/2021/7433186>
- [66] Park, C.W., Oh, S.J., Kim, K.S., Jang, M.C., Kim, I.S., Lee, Y.K., Chung, M.J., Cho, B.H., Seo, S.W., Kim, E.J. (2022). Artificial intelligence-based classification of bone tumors in the proximal femur on plain radiographs: System development and validation. *Plos One*, 17(2): e0264140. <https://doi.org/10.1371/journal.pone.0264140>
- [67] von Schacky, C.E., Wilhelm, N.J., Schäfer, V.S., Leonhardt, Y., Gassert, F.G., Foreman, S.C., Gassert, F.T., Jung, M., Jungmann, P.M., Russe, M.F., Fehrenbach, M. (2021). Multitask deep learning for segmentation and classification of primary bone tumors on radiographs. *Radiology*, 301(2): 398-406. <https://doi.org/10.1148/radiol.2021204531>
- [68] He, Y., Pan, I., Bao, B., Halsey, K., Chang, M., Liu, H., Peng, S., Sebro, R.A., Guan, J., Yi, T. (2020). Deep learning-based classification of primary bone tumors on radiographs: A preliminary study. *EBioMedicine*, 62: 103121. <https://doi.org/10.1016/j.ebiom.2020.103121>
- [69] Ambalkar, S.S., Thorat, S. (2018). Bone tumor detection from MRI images using machine learning. *International Research Journal of Engineering and Technology*, 5(5): 3561-3564. <https://mail.irjet.net/archives/V5/i5/IRJET-V5I5763.pdf>
- [70] Georgeanu, V.A., Mămuleanu, M., Ghiea, S., Selis, D., Teanu, M. (2022). Malignant bone tumors diagnosis using magnetic resonance imaging based on deep learning algorithms. *Medicina*, 58(5): 636. <https://doi.org/10.3390/medicina58050636>
- [71] Ouyang, H., Meng, F., Liu, J., Song, X., Li, Y., Yuan, Y., Wang, C., Lang, N., Tian, S., Yao, M. (2022). Evaluation of Deep Learning-Based automated detection of primary spine tumors on MRI using the turing test. *Frontiers in Oncology*, 12. <https://doi.org/10.3389/fonc.2022.814667>
- [72] Dolatabadi, E., Taati, B., Mihailidis, A. (2017). An automated classification of pathological gait using unobtrusive sensing technology. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(12): 2336-2346. <https://doi.org/10.1109/TNSRE.2017.2736939>
- [73] Teufl, W., Taetz, B., Miezal, M., Dindorf, C., Fröhlich, M., Trinler, U., Hogan, A., Bleser, G. (2021). Automated detection and explainability of pathological gait patterns using a one-class support vector machine trained on inertial measurement unit based gait data. *Clinical Biomechanics*, 89: 105452. <https://doi.org/10.1016/j.clinbiomech.2021.105452>
- [74] Ramirez, E., Wimmer, M., Atzmueller, M. (2019). A computational framework for interpretable anomaly detection and classification of multivariate time series with application to human gait data analysis. In *Artificial Intelligence in Medicine*, Poznan, Poland, pp. 132-147. [https://doi.org/10.1007/978-3-030-37446-4\\_11](https://doi.org/10.1007/978-3-030-37446-4_11)
- [75] Guo, G., Guffey, K., Chen, W., Pergami, P. (2017). Classification of normal and pathological gait in young children based on foot pressure data. *Neuroinformatics*, 15(1): 13-24. <https://doi.org/10.1007/s12021-016-9313-x>
- [76] Lee, S.S., Choi, S.I. (2019). Classification of gait type based on deep learning using various sensors with smart insole. *Sensors*, 19(8): 1757. <https://doi.org/10.3390/s19081757>
- [77] Ramkumar, P.N., Haeberle, H.S., Ramanathan, D., Cantrell, W.A., Navarro, S.M., Mont, M.A., Bloomfield, M., Patterson, B.M. (2019). Remote patient monitoring using mobile health for total knee arthroplasty: Validation of a wearable and machine Learning-Based surveillance platform. *The Journal of Arthroplasty*, 34(10): 2253-2259. <https://doi.org/10.1016/j.arth.2019.05.021>