International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Automatic Diagnosis of Fracture using Deep Learning and External Validation: A Systematic Review and Meta-Analysis

[1]Mr. Irfan Khatik, [2]Dr Sachin Kadam, [3]Dr Milind Gayakwad, [4]Dr Rahul Joshi, [5]Dr Ketan Kotecha

Submitted: 04/12/2023 **Revised:** 12/01/2024 **Accepted:** 27/01/2024

Abstract: Deep learning is a hot area for automatically diagnosing X-rays for bone fractures. Scientists are constantly working on improving clinical practices by exploring new methods. Identifying the fracture, especially hidden, using an automated method is still challenging. Very less external validation on the already studied method is available. This systematic review investigates where current artificial intelligence research stands in assisting radiologists with the correct diagnosis of bone fracture and what are the future directions.

The hybrid approach model is necessary for images collected from the ImageNet Dataset or the Hospital's Radiology department. The processing using the variants of CNN helps in acquiring adequate accuracy for detecting the fracture in the bone in the X-ray image. X-Ray images have been taken, and the output is compared with the pre-trained dataset from ImageNet like InceptionV3, Resnet50, and VGG16.

A systematic review was performed on PubMed and Google Scholar for the studies published between 2019 and 2023. We have included ten articles for the study. These articles are thoroughly analysed and compared for factors like Accuracy, dataset, bone types, etc. External validation is also analysed for each study.

Research in detecting bone fractures through deep learning is continuously increasing. The deep learning model is a good aid in assisting radiologists and clinicians in detecting fractures. Various studies have been performed on bones, but most still need more external validation and heterogeneous data.

***Keywords** – CNN, deep learning, External Validation, fracture*.

1. Introduction

The studies in automatic diagnosis of bone fracture using X-rays and CT (Computerized Tomography) images are increasing. Researchers are trying to explore various algorithms for the accurate detection of fractures applied on various bones. After 2019, there was a significant increase in the articles published on topics, as shown in Fig. 1. .In 2022, the studies highly increased 2022, as shown in Fig.2 (data collected from Google Scholar from 2019 to 2021 and in 2022)

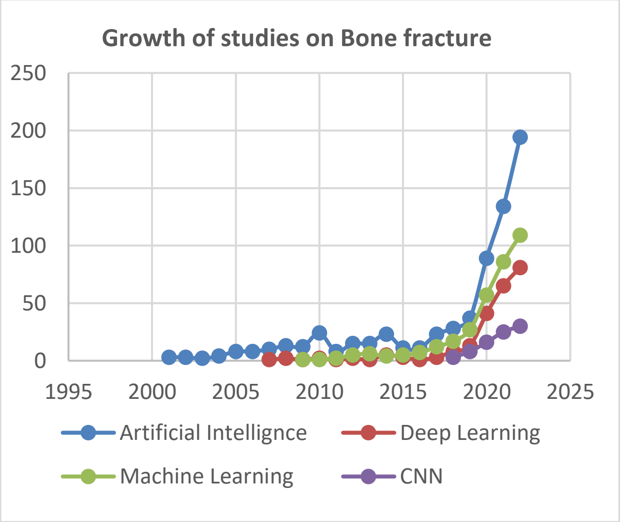

Figure.1 The count of papers on PubMed, when searched using the keywords 'bone fracture' using 'deep learning' or 'machine learning', shows an increase of AI in the automated detection of fractures.

[1]Bharati Vidyapeeth Deemed to be University (Yashwantrao Mohite College, Pune, India) Pune-411038, India
ORCID ID: 0000-0002-6794-5290
[2]Bharati Vidyapeeth Deemed to be University (Institute of Management and Entrepreneurship Development) Pune-411038, India
ORCID ID: 0000-0001-9330-3526
[3]Bharati Vidyapeeth Deemed to be University (College of Engineering) Pune-411043, India
ORCID ID: 0000-0002-0430-2626
[4]Symbiosis Institute of Technology Pune, Symbiosis International (Deemed University), Pune, India
ORCID ID: 0000-0002-5871-890X
[5]Symbiosis Institute of Technology Pune, Symbiosis International (Deemed University), Pune, India
ORCID ID: 0000-0003-2653-3780
** Corresponding Author Email: Irfan.khatik@fergusson.edu*

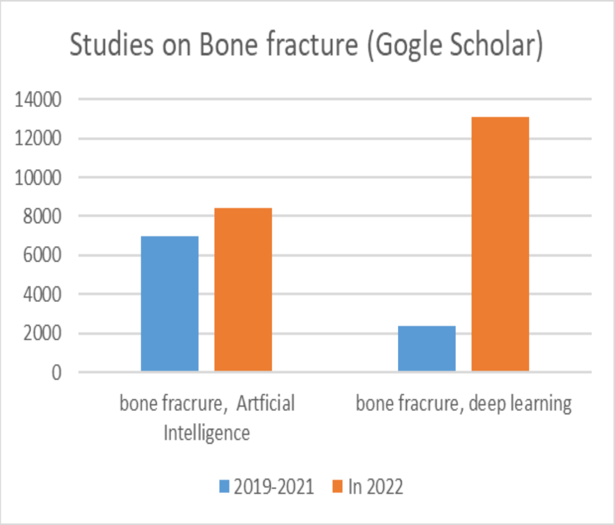

Figure 2 Comparison of studies on Google Scholar using keywords 'Artificial Intelligence, 'CNN', or 'deep learning' with 'bone fracture' from 2019 to 2021 and 2022

External validation is another crucial point to check whether the model is applicable. The number of deep learning models for bone fracture detection and classification is increasing, but models with external validation on a different data set are comparatively shallow. To practise the auto-detection of fractures in clinics, the external validation of the Model is necessary. The model must be validated using external datasets from other hospitals. "The number of externally validated CNNs in orthopaedic trauma for fracture recognition is still scarce" [13]. Fig 3 compares papers on Bone fracture detection using Artificial Intelligence published on PubMed with and without external validation. Of 1326 articles, only 74 (17%) used external validation.

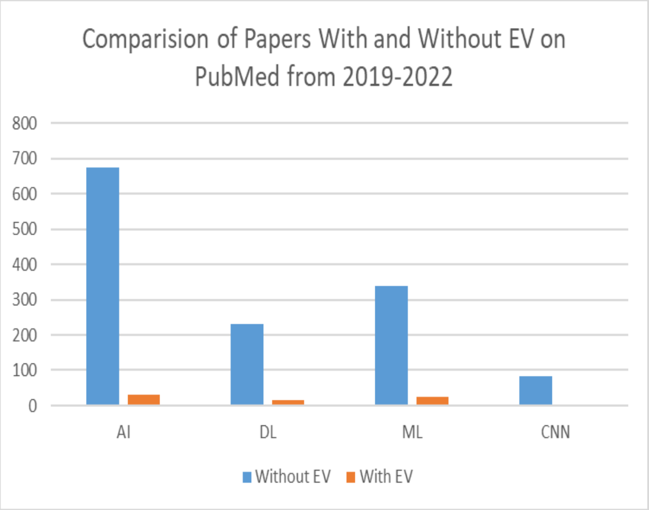

Fig 3 Comparison of studies on bone fracture with and without External Validation (EV)

Now, research using deep learning techniques is increasing in all the fields of radiology. Anis et al. 2020[11] reviewed the deep learning techniques on chest radiographs. Some studies also focused on deep-learning and machine learning-based fracture diagnosis for a particular period. S. Yang [1] etc. reviewed the accuracy of diagnosis using deep learning in orthopedic fractures. Shelmerdine et al. [3] reviewed radiological pediatric fracture assessments using Artificial Intelligence [3]. Various bone fracture detection techniques were reviewed by Khatik [4]. Machine learning and deep learning-based models increase clinicians' diagnostic accuracy, and these models can complement the clinicians and not replace them [7]. Anderson concludes that clinicians' performance will increase when aided by the machine learning system [8].

2. RESEARCH METHODOLOGY

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is used as a basis in this systematic review.

A PubMed database search gave 1326 articles on human bone fracture diagnosis and classification using deep learning. Search is performed using the keyword 'bone fracture' with 'artificial intelligence' or 'deep learning' or 'machine learning'' on PubMed, and a search on Google Scholar with bone fracture using CNN or artificial intelligence or deep-learning or machine-learning gives 19400 results. The PRISMA flow diagram in fig4 reflects the selection of studies for this review.

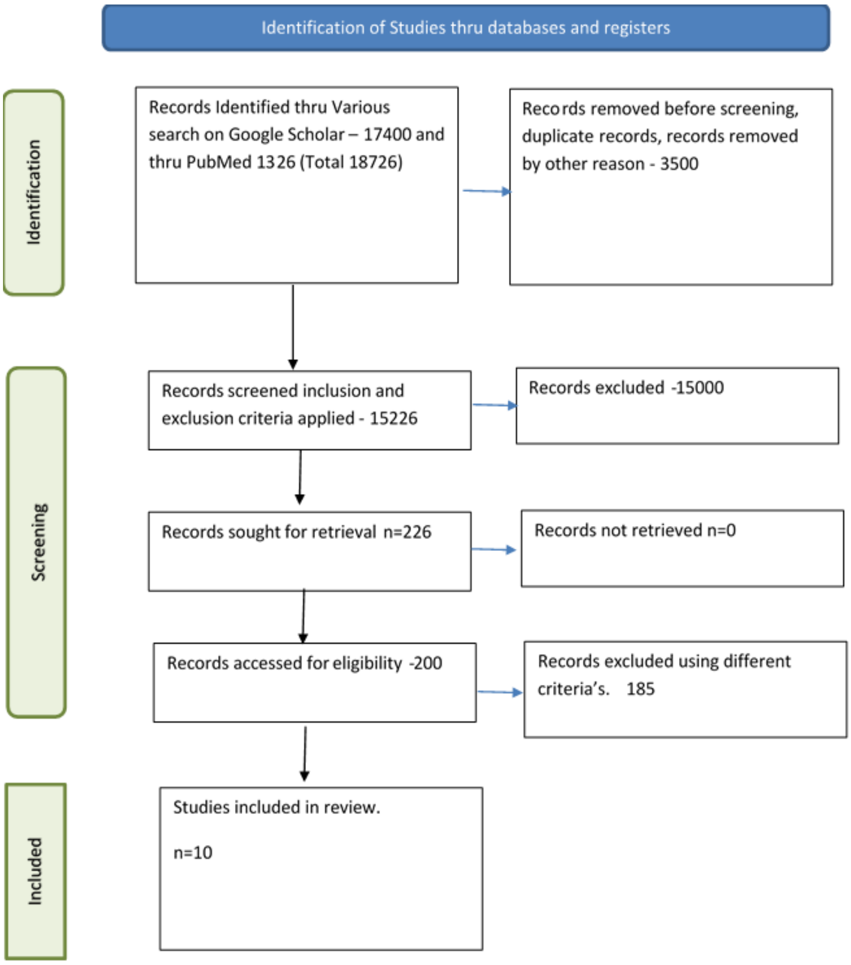
Identification of Studies thru databases and registers
Identification
Records Identified thru Various search on Google Scholar – 17400 and thru PubMed 1326 (Total 18726)
Records removed before screening, duplicate records, records removed by other reason - 3500
Screening
Records screened inclusion and exclusion criteria applied - 15226
Records excluded -15000
Records sought for retrieval n=226
Records not retrieved n=0
Records accessed for eligibility -200
Records excluded using different criteria's. 185
Included
Studies included in review.
n=10

Figure 4. PRISMA flow diagram depicts the articles selected for review

.

3. DISCUSSION

Studies on automatic diagnosis of fractures using deep learning are increasing daily. Our focus is on selected studies as per the above criteria. As per the collected and studied papers, very few studies use external validation based on temporal or geographical data for automated detection. Fracture studies have been performed on various bones, and researchers have applied varied algorithms for the detection.

.

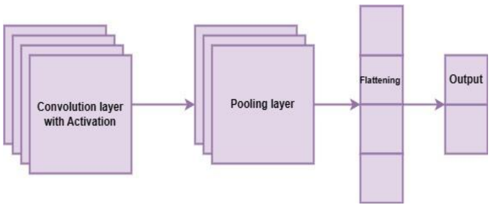

Figure. 5 Convolutional Neural Network in Brief

As mentioned in Fig. 5, the convolutional neural network can be well interpreted using Activation Layer, Pooling Layer, Flattening and Output. The conventional deep learning model may need help overfitting the large feature size. The Convolution technique converts the data into different forms using activation. Max pooling considers only the maximum value from each representation. The flattening and pooling help in feature reduction. The output is then calculated, which is generally the class membership.

Researchers [9] use a faster RCNN deep-learning technique to automatically detect and classify tibia-fibula fractures. The dataset is obtained from a hospital's Diagnostic centre in Islamabad Model is retrained on X-ray (50 images) of Faster RCNN applied on tibia-fibula bones and analysed six classes of bone fracture. The model used 40k steps in training. As evaluation parameters for defining accuracy, the authors used mean-average-precision (map) and kappa coefficient for this study. R-CNN trained using inception v2 networks. The model is trained using the Stochastic Gradient Descent (SGD) method until loss reaches 0.0005%. The authors used a deep–fully–convolutional network (DFCN), an ordinary CNN with a difference for detection and segmentation. In DFCN, another CNN replaced the last fully connected layer. Bounding boxes were used to find the region of fracture.

The researchers used the Kappa coefficient to evaluate the model's overall performance. It took 4000 steps in training, and the total time is 72 hours. Python 3.4 with TensorFlow framework used for training. According to the authors, the accuracy achieved is 97% on the tibia-fibula, and its applicability should be checked on other long bones.

Wu J [6] Worked on rib fracture detection (multiple) and designed a CNN model based on You Only Look Once v3 (YOLOv3). The model is tested on a group of X-rays having 162 fractured and 233 non-fractured. For training, heterogeneous data from four different hospitals is used. In this study, radiographs were labeled according to the doctor's expertise.

The dilated-CNN and long short-term memory (LSTM) algorithm was applied by Tooba Rashid, Zia [4] to diagnose wrist fractures. The author applied the multi-feature extraction approach to classify wrist fracture, tested the model on Augmented and non-augmented data, and achieved a better accuracy of 88% on augmented data against 86% on non-augmented data.

ResNet-18 with convolutional block attention module (CBAM) + + was applied on Hip and pelvic X-rays for detecting Femoral Neck fracture (FNF). X-rays from 2005 to 2018 were used in this study. Evaluation matrices used are under the curve (AUC), accuracy, sensitivity, specificity and Youden index. The important thing about this study is that external validation (EV) is performed for the proposed CNN.

Oakden-Rayner et al. [14] worked on proximal x`femur fracture detection with external validation on X-rays from a hospital emergency department. This study compares the model's performance against five radiologists on a dataset of 200 fracture cases and 200 non-fractures. Performance is good after external validation, but some limitations were observed during testing.

A combined study on adult and pediatric radiographs was performed by Huhtanen JT [15]. E The results of the model were also compared with three radiologists. The average accuracy of the model is 92.8%. External validation is not used for the study, and there is a limited dataset, so this study cannot be generalized.

A 3D Dense Net was used by the researcher Yao L et. [8] to detect rib fracture detection on CT images. A total of 1707 patient's CT images were studied out of which 1507 for training and 100,100 for validation and testing. "The F1-score, precision, recall and NPV values of the model were 0.890, 0.869, 0.913 and 0.969 respectively". The study depicts that artificial intelligence improves the performance of radiologists. Two separate groups of junior and experienced radiologists diagnosed the fracture. The F1-score of the junior had improved from 0.796 to 0.925, and that of experienced radiologists increased from 0.889 to 0.970, respectively. For the dataset, all the chest CT images were collected and annotated by experienced radiologists from a hospital in China. The bone fracture U-Net model, an encoder-decoder architecture based on a Convolution Neural Network, was trained. A 3D Dense Net used for rib fracture classification.

Ukai K [17] also worked on pelvic fracture detection for CT images. Multiple 2D images are used. The researcher has applied DCNN based on the YOLOv3 model. The model was validated on 93 subjects with different orientations with fractures. And 112 subjects without fractures. Modified Faster-RCNN with a rotating bounding box to the long bone fractures was applied by Vironicka[18]. They achieved an accuracy of 0.961. The dataset is obtained from a government hospital in Chennai. Stochastic Gradient Descent (SGD) with a rate of 0.0001 is used to train the Faster R-CNN model. Anttila et al.[19] developed the model using pixel-level annotations of fractures for precise distal radius fracture detection. Adam optimiser was used with a learning rate of 0.001.

The following table-1 shows the summary of the studied papers.

Table 1- Summary of the studied papers

Study	Image type	Bone type	Input Train: Validation: Testing	Model	EV Size	EV	Performance	Year
Yao et al[8]	CT image	Rib fracture	1707 1507:100:100	A 3D DenseNet	-	-	F1-score 0.890 Recall 0.913 Precision 0.869 NPV 0.969	2021

Abbas, Waseem et al. [10]	X-ray	Tibia-fibula fracture	50 images	Faster R-CNN using inception v2	-	-	Accuracy -97% Kappa 97.5%	
Rashid, T.; Zia et. Al.[5]	X-ray	Wrist fracture	3484:358	Dilated CNN-LSTM	-	-	Accuracy -88.24 Precision -87.93 Sensitivity - 92.17 Specificity -82.93 F1-score -90 Kappa -75.7	2023
Wu J, Liu N [6]	X-ray	Chest fracture	918:162	YOLOv3	-	-	Accuracy -85.10 Precision -81.00 Sensitivity - 93.20 Specificity -79.40 AUC -0.92 FROC -91.3%	2023
Femoral Neck Junwon Bae [12]	X-ray	Hip and Pelvic	4189 images 80%;10%:10%	ResNet-18 and a new Conventional network with a block attention module (CBAM) + +	2099	Yes	AU, -0.99 accuracy 0.96 Youden index 0.96, sensitivity 0.96, specificity 0.99+ After EV AUC - 0.97 Accuracy -0.97 Youden index-0.92, sensitivity -0.93, specificit	2021

							y - 0.98	
Oakden-Rayner et. al.[4]	x-ray	Hip, Pelvic	4577	Deep learning model	81	Yes	AUC-0.99 Sensitivity-94 Specificity-99 After EV AUC-0.98	2022
Huhtanen JT[15]	X-ray	Elbow	666:222:213	VGG16, DenseNet201, MobileNet, ResNet152, Inception V3, NASNetLarge,CheXNet.	-	No	AUC- 0.95 Accuracy - 89.8%, precision-86.8%, Sensitivity- 88.8%, Specificity- 90.5%, F1 measure-87.8%, Cohen's kappa-0.80	2022
Ukai . et. al. [17]	CT image	pelvic fracture	A:93 subject B:112 subject	YOLOv3 DCNN	-	NO	AUC - 0.82 recall-0.80 precision - 0.91 Fscore A:0.80,B:0.9 specificity - 0.96	2021
Veronica S et al.[18]	X-ray	long bone fracture	200	Faster R-CNN with	-	No	Accuracy - 0.96,	2023

				rotated bounding box. VGG-16			Sensitivity - 98.6, Specificity -92.7, Precision - 98.6, F1-score - 98.6	
Anttila et al.[19]	X-ray	Wrist fracture	3399:386	Segmentation model based on Unet with 25 layers	-	No	AUC - 0.96, Accuracy -0.91, Sensitivity -0.92, Specificity -0.88.	2022

As per these studies, very few papers use external validation, which will give good accuracy for the given dataset, but accuracy may degrade for an external dataset from another hospital. Some researchers are dividing the dataset for training and validation. However, for a validation set from another source or hospital, a deep learning algorithm degrades the performance. We are suggesting a proposed model for deep learning-based bone fracture detection.

After performing pre-processing, the model should be internally validated, and after that, the model should be used in clinical practice. The Model should be externally validated using real-time data from that hospital. The model must be tested for data collected in the variable period. The following fig. shows the suggested architecture.

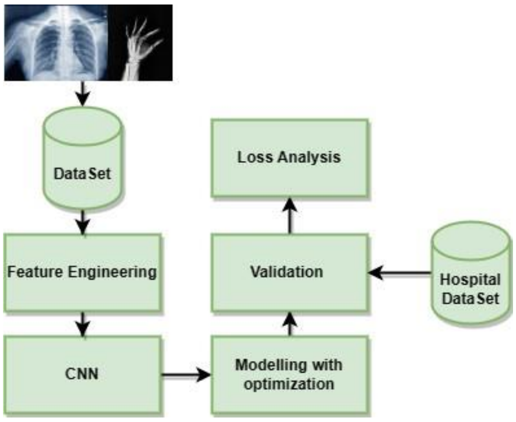

Figure 6 Model with external validation for bone fracture detection

As indicated in Fig.6, the augmented dataset of the X-ray images has been taken, and the output is compared with the pre-trained dataset from ImageNet like InceptionV3, Resnet50, and VGG16.

Acknowledgement

This research was funded by “Research Support Fund of Symbiosis International (Deemed University), Pune, Maharashtra, India”.

4. CONCLUSION

The deep learning-based bone fracture detection will be helpful to physicians in emergency and telemedicine online consultations. It can be helpful to detect hidden, unobvious fractures. To apply artificial intelligence-based bone fracture detection systems in clinical practices, a standard system is needed to assess the performance of the model currently being developed. This study shows that few studies use external validation for bone fracture detection. The models should be tested using the images of different hospitals as external validation. It also shows that external validation increases the accuracy of the mode, making it more applicable.

Funding

This research was funded by “Research Support Fund of Symbiosis International (Deemed University), Pune, Maharashtra, India”.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Yang, S. & Yin, B. & Cao, W. & Feng, C. & Fan, G. & He, S.. (2020). Diagnostic accuracy of deep learning in orthopaedic fractures: a systematic review and meta-analysis. Clinical Radiology. 75. 10.1016/j.crad.2020.05.021.

[2] Bone Jt Open 2021;2-10:879–885

[3] Shelmerdine, Susan & White, Richard & Liu, Hantao & Arthurs, Owen & Sebire, Neil. (2022). Artificial

intelligence for radiological paediatric fracture assessment: a systematic review. Insights into Imaging. 13. 10.1186/s13244-022-01234-3.

[4] Khatik, I. (2017). “A study of various bone fracture detection techniques”. International Journal of Engineering and Computer Science, 6(5), 21418-21423

[5] Rashid, T.; Zia, M.S.; Najam-ur-Rehman; Meraj, T.; Rauf, H.T.; Kadry, S. A Minority Class Balanced Approach Using the DCNN-LSTM Method to Detect Human Wrist Fracture. Life 2023, 13, 133. https://doi.org/10.3390/life13010133

[6] Wu J, Liu N, Li X, Fan Q, Li Z, Shang J, Wang F, Chen B, Shen Y, Cao P, Liu Z, Li M, Qian J, Yang J, Sun Q. Convolutional neural network for detecting rib fractures on chest radiographs: a feasibility study. BMC Med Imaging. 2023 Jan 30;23(1):18. doi: 10.1186/s12880-023-00975-x. PMID: 36717773; PMCID: PMC9885575.

[7] Groot OQ, Bongers MER, Ogink PT, Senders JT, Karhade AV, Bramer JAM, Verlaan JJ, Schwab JH. Does Artificial Intelligence Outperform Natural Intelligence in Interpreting Musculoskeletal Radiological Studies? A Systematic Review. Clin Orthop Relat Res. 2020 Dec;478(12):2751-2764. doi: 10.1097/CORR.0000000000001360. PMID: 32740477; PMCID: PMC7899420.

[8] Yao L, Guan X, Song X, Tan Y, Wang C, Jin C, Chen M, Wang H, Zhang M. Rib fracture detection system based on deep learning. Sci Rep. 2021 Dec 6;11(1):23513. Doi: 10.1038/s41598-021-03002-7. PMID: 34873241; PMCID: PMC8648839.

[9] Anderson PG, Baum GL, Keathley N, Sicular S, Venkatesh S, Sharma A, Daluiski A, Potter H, Hotchkiss R, Lindsey RV, Jones RM. Deep Learning Assistance Closes the Accuracy Gap in Fracture Detection Across Clinician Types. Clin Orthop Relat Res. 2023 Mar 1;481(3):580-588. doi: 10.1097/CORR.0000000000002385. Epub 2022 Sep 9. PMID: 36083847; PMCID: PMC9928835.

[10] Abbas, Waseem & Adnan, Syed & Javid, Dr & Ahmad, Wakeel. (2021). Analysis OF Tibia-Fibula Bone Fracture Using Deep Learning Technique Of X-Ray Images. International Journal for Multiscale Computational Engineering. 19. 10.1615/IntJMultCompEng.2021036137.

[11] Anis, Shazia & Lai, Khin Wee & Chuah, Joon Huang & Ali, Muhammad & Mohafez, Hamidreza & Hadizadeh, Maryam & Ding, Yan & Chao, Ong. (2020). An Overview of Deep Learning Approaches in Chest Radiograph. IEEE Access. 8. 182347 - 182354. 10.1109/ACCESS.2020.3028390.

[12] Bae J, Yu S, Oh J, Kim TH, Chung JH, Byun H, Yoon MS, Ahn C, Lee DK. External Validation of Deep Learning Algorithm for Detecting and Visualizing Femoral Neck Fracture Including Displaced and Non-displaced Fracture on Plain X-ray. J Digit Imaging. 2021 Oct;34(5):1099-1109. doi: 10.1007/s10278-021-00499-2. Epub 2021 Aug 11. PMID: 34379216; PMCID: PMC8554912.

[13] Oliveira E Carmo L, van den Merkhof A, Olczak J, Gordon M, Jutte PC, Jaarsma RL, IJpma FFA, Doornberg JN, Prijs J; Machine Learning Consortium. There are increasing convolutional neural networks for fracture recognition and classification in orthopaedics: are these externally validated and ready for clinical application? Bone Jt Open. 2021 Oct;2(10):879-885. doi: 10.1302/2633-1462.210.BJO-2021-0133. PMID: 34669518; PMCID: PMC8558452.

[14] Oakden-Rayner L, Gale W, Bonham TA, Lungren MP, Carneiro G, Bradley AP, Palmer LJ. Validation and algorithmic audit of a deep learning system for detecting proximal femoral fractures in patients in the emergency department: a diagnostic accuracy study. Lancet Digit Health. 2022 May;4(5):e351-e358. doi: 10.1016/S2589-7500(22)00004-8. Epub 2022 Apr 5. PMID: 35396184.

[15] Huhtanen JT, Nyman M, Doncenco D, Hamedian M, Kawalya D, Salminen L, Sequeiros RB, Koskinen SK, Pudas TK, Kajander S, Niemi P, Hirvonen J, Aronen HJ, Jafaritadi M. Deep learning accurately classifies elbow joint effusion in adult and pediatric radiographs. Sci Rep. 2022 Jul 12;12(1):11803. doi: 10.1038/s41598-022-16154-x. PMID: 35821056; PMCID: PMC9276721.

[16] Ukai K, Rahman R, Yagi N, Hayashi K, Maruo A, Muratsu H, Kobashi S. Detecting pelvic fracture on 3D-CT using deep convolutional neural networks with multi-orientated slab images. Sci Rep. 2021 Jun 3;11(1):11716. doi: 10.1038/s41598-021-91144-z. PMID: 34083655; PMCID: PMC8175387.

[17] Veronica S, Sathiaseelan JGR (2023) Classification of Long-Bone Fractures Using Modified Faster RCNN for X-Ray Images. Indian Journal of Science and Technology 16(1): 56-65. https://doi.org/ 10.17485/IJST/v16i1.1690

[18] Anttila TT, Karjalainen TV, Mäkelä TO, Waris EM, Lindfors NC, Leminen MM, Ryhänen JO. Detecting Distal Radius Fractures Using a Segmentation-Based Deep Learning Model. J Digit Imaging. 2023 Apr;36(2):679-687. doi: 10.1007/s10278-022-00741-5. Epub 2022 Dec 21. PMID: 36542269; PMCID: PMC10039