Computer Vision for Healthcare: Detecting Fractures in Hand X-Rays

C V Sree Pranavi
Department of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4cse22216@bl.students.amrita.edu

C Madhuja
Department of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4cse22212@bl.students.amrita.edu

Dr. Rimjhim Padam Singh
Department of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
ps$_r$$imjhim$@$blr.amrita.edu$

***Abstract*—It is difficult for a doctor to spot fractures in fingers, hands, and wrists on X-rays since it needs a lot of care and can still lead to incorrect identification. Because doctors are often stressed with the many patients to see due to a lack of qualified staff, relying on automation in diagnosis brings many benefits. It relies on deep learning to check if there are fractures visible in musculoskeletal images viewed by X-ray. The purpose of this method is to handle problems such as resolution of images, patient variations in size, and the lack of labelled data. Initially, light theoretical models, including custom convolutional network and Mobile Network Version Two, are used. From there, Inception Version Three, Densely Connected Network, Neural Architecture Search Network, Vision Transformer, and Extreme Inception Network are introduced. To enhance how the network can predict, Squeeze and Excitation, Convolutional Block Attention Mechanism, self-attention, and Global Context Attention attention tools are added to version B3 of the Efficient Network. From the tested models, Efficient Network B3 Version B3 with Global Context Attention was ranked first, having an accuracy rate of 86%. Furthermore, the model performed accurately, as seen by high precision, recall, and F1-score, and experienced a low loss on the test data. Deep learning makes it possible for the system to give reliable and appropriate help in detecting fractures in hospitals and improves the work and health of patients.**

***Index Terms*—Fracture Detection, X-ray Imaging, Deep Learning, Attention Mechanisms, Diagnostic Automation**

I. Introduction

Reading X-rays of fingers, hands, and wrists can be hard as the bones and joints in these areas are complex and the changes caused by illnesses are often very small. It is usually done by trained radiologists who pay close attention and have lots of skill. However, sometimes people's tiredness and variant ways of interpreting things may lead to diagnosis mistakes. As a result, the growing number of medical imaging examinations means that there are not enough trained radiologists to meet the demand for diagnoses. For this reason, accurate and fast automated diagnostic tools are needed more than ever before.

Building such systems raises certain challenges of its own. High-quality, annotated datasets are not easy to find due to the intensive and expert-based process of labeling medical images. The inconsistent quality of radiologic images, patients' movement, and differences in age and gender are some of the other challenges in simulation. Spotting little abnormalities like micro-fractures requires a lot of attention and advanced tools. The model should be effective for different populations and types of imaging devices, but it should not get too complacent and memorize the patterns in its training data. Also, for clinical deployment to be successful, healthcare AI must fit smoothly into existing hospital routines and ensure all data privacy and compliance rules are followed.

Before, people would look at images and recognize their contents, or machine learning would be done using features that were written by hand. It is difficult for them to address complex patterns, so they are mainly used for small projects. Other radiographic systems depend on applying rules to generate indexes. Still, since they are not able to adapt and require constant expert presence, they may not suit hospitals with a lot of work to do each day.

Deep learning is deployed to check if X-ray films from hand, wrist, and finger reveal the presence of fractures. Neural networks allow the AI to analyze images and find essential patterns, allowing medical pictures to be checked efficiently and accurately. The aim is to develop a system that can recognize fractures easily and solve certain problems encountered by doctors while analyzing X-ray images.

The rest of the paper is organised as follows: Section II looks at what other studies have found, Section III talks about how this research was done, Section IV covers the results and Section V give a summary and future directions.

II. Literature Review

Though using Inception-v3 and Faster R-CNN makes detecting fractures easier and more accurate, they do face prob-

lems due to data problems and different views from different doctors [1]. A mix of ConvNet, ResNet and DenseNet worked better than other methods and human experts on the MURA dataset, reaching an AUC of 0.93, a precision of 0.93 and a recall of 0.81 in three regions. It turns out that the results for the humerus region were excellent, with a Cohen's kappa of 0.85, though the study showed that some conditions may not be detected [2].

Using GNG clustering and updating the VGG model leads to a better performance in bone X-ray image classification [3]. While it works very well, it takes significant computer power and a lot of training data to operate. A CNN-based approach makes it easier to find abnormalities and is more interpretable [4]. Yet, setting up deep learning models is costly, not easy, and requires plenty of time, making them less accessible. They point out that there are obstacles for clinical use in real-life settings.

The analysis relies on a combination of Support Vector Machines, Visual Geometry Group-16, and Xception Version 3 neural networks to identify and spot abnormalities in X-rays [5]. At the first level, X-rays are organized into seven body regions, and Visual Geometry Group-16 was shown to be correct almost 95% of the time. Local Binary Patterns and Gray-Level Co-occurrence Matrix are used on the second level to detect abnormalities, but the accuracy was 66.33%. Using Adaptive Residual Network Split-Attention along with Attention-Convolution Bidirectional Feature Pyramid Network, they managed to get 68.4% precision in just 122 milliseconds [6]. But, depending on private data and doing only as much optimization as possible makes it hard to use the research widely.

The move from traditional machine learning to deep learning in bone analysis and spotting irregularities has allowed for higher accuracy, with early models still achieving 85–94%, but needing manual action to identify features. However, reaching nearly 97 percent accuracy, ResNeXt50 was shown to be sensitive when the data set was limited in samples. Although information was increased, learning to perform outside the training environment is still a problem. With the annotated abnormalities and lesions in the MURA-objects dataset, it was possible to use Faster Region-Based Convolutional Neural Network and You Only Look Once Version 3 for better evaluation [8].

In a study when deep learning models were tested using the MURA dataset, DenseNet169 turned out to be the most accurate, with a 87.88% training accuracy and 79.20% validation accuracy, suggesting that its ability to generalize could still be improved [9]. Doctors used a model based on Xception and an SVM classifier to detect the type of bone and point out its irregular shape [10]. Both single-view and multi-view ways of collecting data improved the accuracy of diagnoses. Adding the Support Vector Machine layer helped the model be more robust, but this increased calculation time. This approach is not well-suited for use in situations where resources are constrained.

Ensemble classification of wrist images has been put forward, with UNet used to segment tags and make a diagnosis using group voting, improving accuracy by 1.5–4.5% compared to manual and other automatic methods [11]. Even though it can deal with various types of data, it is not accurate enough and needs much more labeled data to train UNet. Working with deep learning and Capsule Networks, the MSDNet approach reaches a Class AMS result of 82.69% [12]. They aid in automatically detecting diseases and reduce the amount of work that falls on radiologists. However, difficulties persist in the collection of data, designing the models correctly, and the difficulty of explaining how they work.

To find bone abnormalities in the past, only basic methods like edge detection were used, but they were not strong enough. Support vector machines and k-nearest neighbors give a wide range of applications, but they are not always accurate. While deep learning, in particular convolutional neural networks, is good at recognizing images, it has difficulties modeling complicated relationships. By including lesion-guided regions, the proposed LGAG-Net was able to reach 87.81% accuracy, but its use needs a lot of computing resources and testing on more data sets [13]. Although this method can be automated and offers great accuracy, it has the problem of overfitting and high computation time [14].

The models that are tested in the study include Inception-v3, VGG16, DenseNet-121/169, and ResNet-50 on datasets named MURA and Kaggle for detecting wrist and elbow bones. Data sets can affect models, as shown by DenseNet-169 outperforming on MURA and Inception-v3 doing better on Kaggle [15]. In another trial, convolutional neural networks made it possible to improve how well and fast finger X-rays could be analyzed [16]. While they are very useful, using these models demands strong computers and makes them hard to understand, so they are rarely used in clinics.

Classical feature extraction uses the judgments from experts, and there is a risk of making mistakes. By contrast, deep learning is more accurate but comes with greater computing demands. Networks designed for mobile devices are simple and portable, but adding different types of data makes it more accurate but also more complicated to use [17]. DenseNet, ResNet, transfer learning, and ensemble learning allow shoulder abnormality detection with high accuracy, but issues with overfitting, how features are extracted, and the hardness of training processes still exist [18].

A deep learning model was designed in this study and found to accurately divide musculoskeletal X-rays into eleven groups, detecting approximately 89 percent of abnormalities, while maintaining a 97.37 percent accuracy, with good ROC-AUC and specificity. Even so, because the sensitivity is only 0.86, the tool may not discover every problem and it still requires further improvement and a lot of computer resources [19]. A different method brings together explainable AI and convolutional neural networks to help doctors decide on diagnoses, but it requires more work to adjust and could be quite expensive before being used on a large scale [20].

III. METHODOLOGY

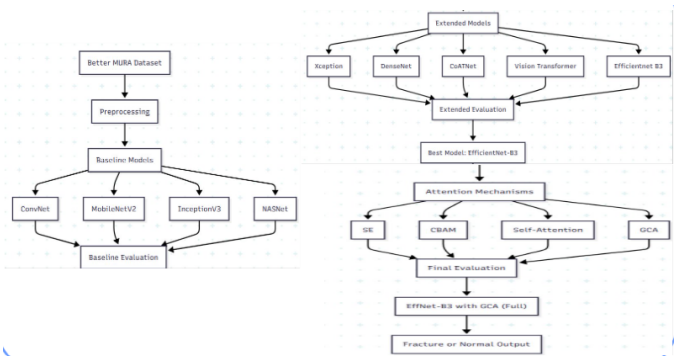

Fig. 1. Architecture Flow Diagram

Fig. 1. shows the workflow of the suggested methodology, presenting the entire sequence of the project from initiation to completion. It defines each step, starting with preprocessing musculoskeletal X-ray images from the Better MURA dataset, followed by model training on different deep learning architectures. The figure also shows the incorporation of attention mechanisms into the top-performing model, and ends with performance testing and ultimate classification into fractured or normal cases.

A. Dataset Preparation

The Better MURA dataset has an initial large collection of musculoskeletal radiographs with both original and augmented images. But in order to provide the model with unique and non-resourced data, all augmented images were discarded, and resampling (undersampling/oversampling) is not used, but stratified splitting (60% train, 20% validation, 20% test) provides balanced class distribution, keeping only the original X-ray scans.

After preprocessing, the dataset was segregated into two categories based on the presence or absence of fractures:

- Class (Fractured): Holds 4,639 X-ray images classified as having a fracture.
- Class (Non-Fractured): Holds 6004 X-ray images classified as normal (no fractures).

This labeled dataset allows for a well-defined split between classes, which enables effective training and testing of deep learning models for classifying fractures.

B. Baseline CNN Architectures

1) Model Selection: The following baseline CNN architectures were taken into consideration for comparison in the study: ConvNet, InceptionV3, DenseNet, Xception, NASNet, and MobileNetV2. They illustrate many different breakthroughs in architecture and specific performance plots are made for image classification tasks.

- ConvNet: This is the most fundamental example of a convolutional neural network used as a foundational baseline. It helps establish a refer-Key point to assess how much the performance has improved offered by more complex architectures.
- InceptionV3: Chosen because it has a unique design that uses multiple convolutional paths each using a different-sized filter. InceptionV3 captures multiscale features effectively while maintaining computational efficiency.
- DenseNet: This architecture features densely connected layers which take input from the before layers. It increases the flow of gradient and helps to improve features. Using it several times, making it precise and reliable, especially on complex datasets.
- Xception: Based on depthwise separable convolutions, Xception model is used due to its strong capabilities in reducing the training required because of the number of parameters and the computation it takes in compromising model accuracy.
- NASNet: Developed through Neural Architecture Search, NASNet was designed to be the best and most efficient model. Using AI, images can be processed for both high accuracy and efficiency. It is considered as a leading benchmark for today's deep learning.
- MobileNetV2: Intended for use in devices and in mobile computers. The design uses inverted residual connections and depth-wise separable convolutions. It is chosen to demonstrate how small models are just as likely to match the speed of their opponents in resource-constrained environments.

2) Preprocessing: All images in the dataset were preprocessed the same way to make sure results are consistent for all models. All images were resized to 299×299 in order to comply with the input rules of Xception and InceptionV3 deep learning models. The pixels were changed so that their values range from 0 to 1 by dividing by 255.0, allowing the model to train more quickly. As both the data and its labels are clean, no tag removal or data augmentation is done here, so all baseline CNN models work with the same consistent input.

3) Model Architecture: All baseline models were implemented using a transfer learning method, which uses a pre-trained CNN (DenseNet121, Xception, InceptionV3, MobileNetV2, NASNet, and ConvNet) as the fundamental features extractors, trained initially on ImageNet. The top classification layers were removed and the base was frozen to retain pre-trained weights. On top of each base model, a custom classification head was added, consisting of:

- A GlobalAveragePooling2D layer
- A layer using 128 ReLUs.
- A Dropout layer is added with rate 0.5 to avoid overfitting.
- A layer having 64 neurons all using ReLU activation
- A Dense layer made up of 16 ReLU neurons
- A last dense layer contains one unit and uses the sigmoid activation function for binary classification

C. Transformer and Attention-Based Models

1) Model Selection: Self-attention and its strength in finding important global information made Vision Transformer

(ViT) the right choice for this application. A custom CNN was choosen because it learns locally and trains very quickly, making it possible to evaluate both types of models.

2) Preprocessing: For images used in ViT, preprocessing consisted of resizing them to 224×224 and then using a processor to normalize their variances. On the other hand, CNN used a straightforward preprocessing step-shrinking the images to 224×224 and adjusting their pixel values to the [0,1] range by dividing by 255.0.

3) Model Architecture: The model used is google/vit-base-patch16-224-in21k, which is a transformer-based design that preprocesses the image with 16×16 patches, flattens them linearly, adds positional information, then applies several self-attention layers and finishes up with a binary classification head. It is pretrained on ImageNet-21k. The CNN model uses TensorFlow and includes many Conv2D and DepthwiseConv2D layers with BatchNormalization and ReLU, finally ending with Dropout layers for preventing overfitting. After GlobalAveragePooling2D, dense layers and a sigmoid output layer for categorization are the last parts of the network. The CNN model used here is freshly trained and not based on pretrained data.

D. EfficientNet-B3 Variants

1) Model Selection: With the analysis of both conventional CNNs and transformer models done, several EfficientNet-B3 models were brought into focus. They were considered because they offer a good ratio of accuracy to how much they cost to use. To move forward, the baseline model chosen was EfficientNet-B3, which is based on compound scaling. Progressively, Self-Attention, CBAM, SE, and GCA were added to the EfficientNet-B3 backbone to increase the capabilities of feature extraction.

- EfficientNet-B3: It is the main model that handles the balance between depth, width, and resolution of the input.
- EfficientNet-B3 + Self-Attention: Allows attention to be given to the most informative parts.
- EfficientNet-B3 + CBAM: It sequentially uses channel and spatial attention to focus on relelevant aspects of the features.
- EfficientNet-B3 + SE: Uses recalibration of features in each channel to bring out more information.
- EfficientNet-B3 + GCA: Uses global context modeling to find the connections between scenes.

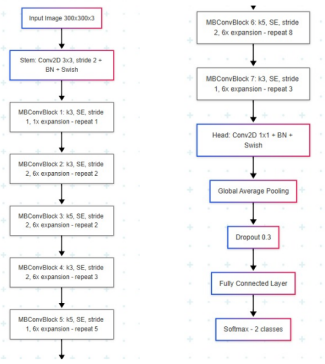

Fig. 2. Architecture of Efficientnet-B3

As shown in Fig.2. an overview of building EfficientNet-B3 begins with an image of size 300×300×3. The Stem section begins with a regular 3×3 convolution which is then normalised and processed by Swish. MBConv blocks now appear which may possess squeeze-and-excitation (SE) modules and allow the kernel size, stride length and expansion to be modified. The way the blocks are repeated changes the size and shape, the process goes from one to eight blocks. At the end of MBConv, the Head layer has a layer consisting entirely of a 1×1 convolution, followed by batch normalization and then Swish. Following that, the spatial sections are merged by using global average pooling and applying dropout with p=0.3 to decrease overfitting. A Softmax layer completes the network with only one fully connected layer, so to produce the probabilities for the two classes.

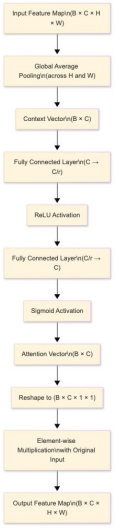

Fig. 3. Architecture of GCA

As shown in Fig.3. to allow the convolutional neural network to learn features from related parts of an image, the Global Context Attention module was added. The initial process is to design the feature map with B × C × H × W, where these numbers represent items, channels and heights and widths. An

aggregated context vector is created by adding the height and width values from the images, after which both the new and original images have the same B and C. Every layer multiplied by a vector with neurons is reduced by 'r' and both the ReLU and Sigmoid functions are run on the vector. In the same way as B × C for the linear example, B × C × 1 × 1 is applied for the convolution and blend it with the first feature map. The model is able to remove irrelevant parts of the input by using attentional weights during the operation.

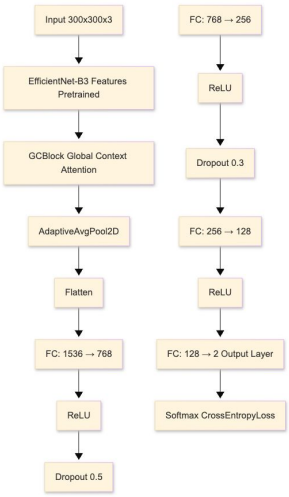

Fig. 4. Efficientnet-b3 + GCA

Fig.4. shows the architecture flow of the proposed model where A Global Context Attention module is introduced after using EfficientNet-B3 for extracting features, followed by multiple classification heads to determine the final binary classification outcome.

1) Preprocessing: All images are processed to obtain a 300×300×3 dimension to be compatible with the input size of EfficientNet-B3. A variety of inputs are supported by applying conventional data preprocessing techniques such as normalization and enabling the model to be effective with diverse types of input data.
2) EfficientNet-B3 Feature Extractor: The EfficientNet-B3 network is responsible for carrying out the vital task of extracting features from the input data. All network parameters are scaled uniformly using a specific set of scaling variables. Scaling all parameters across the model together leads to higher performance while maintaining efficiency.The core of the model is formed by alternate MBConv and ImageNet-inspired SE blocksthroughout the architecture. Adaptable channel-wise modifications make SE modules a valuable addition to the model's ability to represent complex relationships in the extracted features. The EfficientNet-B3 model generates a 1536-channel feature map that captures a wide range of detailed information at various scales.
3) Global Context Attention (GCA) Module: A Global Context Attention (GCA) module is added to the output of EfficientNet-B3 to improve the model's ability to emphasize those parts of the image that are most relevant to the diagnosis process. It aims to highlight and integrate the knowledge present in various parts of the image. The GCA module carries out two main objectives:
   - Context Modeling: Uses a soft attention mechanism to collect and combine all the semantically important pieces of information in the image and giving an overview of how they relate to each other in space.
   - Feature Re-weighting: Reconfigures the importance of features by bringing into account the whole context of the image in order to give value to regions that matter and suppress less important parts.
4) Classification Head: Then, the updated feature map is processed through a series of classification stages outlined by:
   - Adaptive Average Pooling reduces the spatial dimensions and preserves important global information within the feature vector.
   - The flattened feature representation then goes through several fully connected layers containing non-linear activation functions and dropout to reduce the risk of overfitting.
     FC Layer: 1536 → 768 → ReLU → Dropout (p=0.5) FC Layer: 768 → 256 → ReLU → Dropout (p=0.3) FC Layer: 256 → 128 → ReLU
   - A fully connected layer with two output units together with Softmax activation provides the probabilities for binary classification.

With regards to all the variants, EfficientNet-B3 + GCA performed best, reaching an accuracy level of 86%, which points to the benefits of attention mechanisms on model performance.

E. Training Details

The data was trained using Adam optimizer (learning_rate = 0.0001), with binary cross-entropy loss and for 50 rounds on batches of 32 data points. The choice of the epoch and validation step sizes was made automatically using the information given by the dataset. To watch progress during training, validation loss and accuracy were observed. With the training finished, outcomes from the models on the test set were checked, any prediction over 0.5 was given a positive score.

F. Evaluation Metrics

For evaluating the performance of each model, metrics like Accuracy, Loss, Precision, Recall, and F1-score were used. Accuracy counts the correct predictions, and loss represents the mistake a model makes as it trains and validates. Both precision and recall help assess how well the model can discover actual positive cases and also how sensitive it is to them. F1-score allows to easily manage precision and recall, making it most helpful for datasets with uneven classes.

Model	Accuracy (%)
conVNet	71
InceptionV3	76
DenseNet	78
Xception	78
NASNet	78
MobileNetV2	77
Vision Transformer	84
CoATNet	67
EfficientNet-B3	84
Efficientnet-b3 with CBAM	84
Efficientnet-b3 with selfattention	85
EfficientNet-B3 with SE	85
Efficientnet-b3 with GCA	86

TABLE I
MODELS COMPARISON TABLE

IV. RESULTS

Table 1 summarizes all the models that were used in this research work and their respective accuracy percentages. The table demonstrates the evolution from light baseline models to sophisticated architectures, highlighting the performance gain attained at each experiment stage.

Model	Acc. (%)	Loss	Prec.	Recall	F1	Sup.
EffNet-B3	84	0.37	0.84	0.81	0.82	4081
B3 + CBAM	84	0.24	0.87	0.80	0.81	4081
B3 + SelfAttn	85	0.17	0.86	0.82	0.84	4081
B3 + SE	85	0.26	0.87	0.81	0.82	4081
B3 + GCA	86	0.25	0.86	0.83	0.84	4081

TABLE II
EFFICIENTNET-B3 VARIANTS COMPARISON TABLE

Table II shows the performance of various EfficientNet-B3 models evaluated on test accuracy, test loss, precision, recall, F1-score and support. Applying GCA to EfficientNet-B3 achieved the highest F1-score of 0.84 showing that GCA methods enhance the model's ability to strike a balance between accuracy and recall. Models based on EfficientNet-B3 architecture produced similar levels of accuracy, but adding Global Context Attention improved generalization and reduced the test loss and recall error.

Model	Accuracy
ConvNet + ResNet + DenseNet [2]	87%
CNN [4]	68%
DenseNet169 [9]	87.88%
LGAG-Net [13]	87.81%
Proposed Model	89%

TABLE III
STATE-OF-THE-ART MODELS

V. CONCLUSION AND FUTURE SCOPE

The work involved trying out both convolutional and transformer models for classifying binary images, ending up with EfficientNet-B3 models that included various forms of attention. The F1-score shows that using GCA with EfficientNet-B3 improved classification accuracy to 86%, proving the usefulness of context refinement. Using transfer learning and attention modules improved the model's ability to pay attention to the important parts of the images, which helped it learn and work better across many situations. On the whole, these findings support the belief that joining EfficientNet's light architecture with attention has a positive effect on performance under constraints.

This work provides a solid basis for fracture detection with deep learning but there are several aspects to consider for future research. Domain-specific clinical data like patient history or demographics can further increase diagnostic accuracy and provide contextual background. Furthermore, the dataset could be expanded to cover other anatomical regions in the human body. Including varied imaging conditions like poor lighting and quality can help in implementing the model in different scenarios. Real-time deployment via light edge-compatible models may extend this solution to rural and low-resource healthcare facilities, where there is minimal radiologist access. Lastly, implementing this system into hospital information systems and testing its performance on live clinical workflows would be a key step towards large-scale implementation.

REFERENCES

[1] Guan, B., Zhang, G., Yao, J., Wang, X. and Wang, M., 2020. Arm fracture detection in X-rays based on improved deep convolutional neural network. Computers & Electrical Engineering, 81, p.106530.

[2] He, M., Wang, X. and Zhao, Y., 2021. A calibrated deep learning ensemble for abnormality detection in musculoskeletal radiographs. Scientific Reports, 11(1), p.9097.

[3] El-Saadawy, H., Tantawi, M., Shedeed, H.A. and Tolba, M.F., 2021. A Hybrid Two-Stage GNG–Modified VGG Method for Bone X-rays Classification and Abnormality Detection. IEEE Access, 9, pp.76649-76661.

[4] Huynh, H.X., Nguyen, H.B.T., Phan, C.A. and Nguyen, H.T., 2021. Abnormality Bone Detection in X-Ray Images Using Convolutional Neural Network. In Context-Aware Systems and Applications, and Nature of Computation and Communication: 9th EAI International Conference, ICCASA 2020, and 6th EAI International Conference, ICTCC 2020, Thai Nguyen, Vietnam, November 26–27, 2020, Proceedings 9 (pp. 31-43). Springer International Publishing.

[5] Kalivarapu, P., George Rajan, R.F. and Niranjana Krupa, B., 2021. Intelligent Analysis of X-Ray Images for Detecting Bone Abnormality in Upper Extremities. In Proceedings of International Conference on Communication, Circuits, and Systems: IC3S 2020 (pp. 233-239). Springer Singapore.

[6] Lu, S., Wang, S. and Wang, G., 2022. Automated universal fractures detection in X-ray images based on deep learning approach. Multimedia Tools and Applications, 81(30), pp.44487-44503.

[7] Yao, J., Guo, Z. and Yu, W., 2022. Enhanced deep residual network for bone classification and abnormality detection. Medical Physics, 49(11), pp.6914-6929.

[8] Shao, Y. and Wang, X., 2022. MURA-objects: a radioactive bone imaging lesion detection dataset. Machine Vision and Applications, 33(6), p.96.

[9] Rath, M., Reddy, P.S.D. and Singh, S.K., 2022. Deep Convolutional Neural Networks (CNNs) to Detect Abnormality in Musculoskeletal Radiographs. In Second International Conference on Image Processing and Capsule Networks: ICIPCN 2021 2 (pp. 107-117). Springer International Publishing.

[10] Shubhangi, D.C., Gadgay, B. and Waheed, M.A., 2022, November. Medical X-Rays Categorization and Irregularity Recognition of Bone and Diagnosis of Bone disorders Based on Xception Model. In 2022 International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS) (pp. 58-63). IEEE.

[11] Khan, S., Arshad, F., Zulfiqar, M., Khan, M.A. and Memon, S.A., 2022. Ensemble learning-based abnormality diagnosis in wrist skeleton radiographs using densenet variants voting: 10.48129/kjs. splml. 19477. Kuwait Journal of Science.

[12] Karthik, K. and Sowmya Kamath, S., 2023. MSDNet: A deep neural ensemble model for abnormality detection and classification of plain radiographs.Journal of Ambient Intelligence and Humanized Computing, 14(12), pp.16099-16113.

[13] Liao, Y., Li, X. and Peng, C., 2023. LGAG-Net: Lesion-Guided Adaptive Graph Network for Bone Abnormality Detection From Musculoskeletal Radiograph. IEEE Access.

[14] Karanam, S.R., Srinivas, Y. and Chakravarty, S., 2023. A supervised approach to musculoskeletal imaging fracture detection and classification using deep learning algorithms. Computer Assisted Methods in Engineering and Science, 30(3), pp.369-385.

[15] Ismail, M.A., Mohammed, Y., Gamal, N., Hatem, M., Ibrahim, S., Abbas, E.H., Mohsen, M., Nasser, M. and Ebied, H.M., 2023, November. Analysis of Bones Abnormality Detection and Segmentation. In 2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS) (pp. 592-597). IEEE.

[16] Kumar, R., K, S.D. and Mohapatra, D.P., 2024. Assessing radiographic findings on finger X-rays using an enhanced deep learning approach. International Journal of Information Technology, 16(7), pp.4279-4288.

[17] Zeng, Z., Song, C., Liu, Q., Yi, S. and Zhu, Y., 2024. Diagnosis of musculoskeletal abnormalities based on improved lightweight network for multiple model fusion. Mathematical Biosciences and Engineering, 21(1), pp.582-601.

[18] Alzubaidi, L., Salhi, A., A. Fadhel, M., Bai, J., Hollman, F., Italia, K., Pareyon, R., Albahri, A.S., Ouyang, C., Santamaría, J. and Cutbush, K., 2024. Trustworthy deep learning framework for the detection of abnormalities in X-ray shoulder images. Plos one, 19(3), p.e0299545.

[19] Singh, G., Kumar, P. and Anand, D., 2024. Hybrid Deep Learning Model for Classification and Prediction of Abnormalities in Upper and Lower Extremities of Musculoskeletal Radiographs. SN Computer Science, 6(1), p.32.

[20] Wu, Y., Fong, S. and Liu, L., 2025. Enhancing explainability in medical image classification and analyzing osteonecrosis X-ray images using shadow learner system. Applied Intelligence, 55, p.137.