X-Fracture: AI Fracture Detection

Student Name: Ahmed Sami Baabdullah | **Student ID:** 201811980
Supervisor: Dr. Muzammil Behzad [muzammil.behzad@kfupm.edu.sa]
King Fahd University of Petroleum and Minerals, Saudi Arabia

Abstract

Bone fractures are a common medical condition that needs to be diagnosed quickly and accurately in order to be treated effectively. Using radiological images, this project investigates the use of deep learning, part of AI, for the binary classification problem of bone fracture detection. Building on pre-existing convolutional neural networks (CNNs), the model uses architectural improvements to increase classification accuracy while cutting down on training time and cost. The outcomes show significant gains over the use of machine learning (ML) or basic Deep Learning(DL) separately.

1. Introduction

Fractures, often caused by trauma or medical conditions, require early and precise detection for effective treatment and patient recovery. X-ray imaging has been the primary diagnostic method, but manual diagnosis is time-consuming, human error-prone, and relies on medical professionals' skill. Recent advancements in AI are viewed as a promising solution to improve diagnostic efficiency and accuracy in bone fracture detection.

AI fracture detection models face challenges in complex cases like hairline fractures and overlapping bones, and performance can be impacted by patient demographics, image quality, and data accessibility. These issues highlight the need for ongoing advancements in AI detection systems. Enhancing existing models with advanced DL architectures, better data preprocessing methods, and diverse datasets can significantly improve accuracy and reliability, ultimately impacting healthcare workers point of view in AI.

1.1 Problem Statement

Despite advancements in medical imaging, misdiagnosis of fractures remains a significant issue. Traditional methods lack the consistency required for high-accuracy detection. The aim is to improve fracture detection by using a two-stage training approach, which involves first extracting features and then finding best hyper parameters on ML. This directly solves the shortcomings of current methods and aims to provide a scalable solution for fracture detection in healthcare.

1.2 Objectives

- To develop an effective bone fracture detection system based on radiographic images using a transfer learning-based approach.
- To design and implement the MobLG-Net model for robust feature extraction and classification.
- To evaluate the performance of MobLG-Net compared to classical CNN-based methods using large, diverse datasets.

- To validate the generalization capability of the proposed model through cross-validation and performance metrics.

1.3 Scope of Study

To focus on utilizing CNN model for detecting fractures in radiological images and it consider a range of fracture types, including femoral, wrist, and vertebral fractures, by utilizing publicly available datasets like Kaggle,roboflow.

2. Literature Review

Recent studies on fracture detection using radiographic images have primarily used deep learning and machine learning techniques to overcome limitations of traditional manual diagnosis. Guan et al. achieved an average precision of 62.04% with 4,000 arm x-ray images, while Kim and MacKinnon demonstrated that transfer learning with deep Convolutional Neural Networks could improve detection accuracy by up to 95%.

DL, which is result of advancement in AI recently particularly CNN has become a powerful tool in various fields for like healthcare. CNN can accurately classify and locate fractures, with models like VGG16, GoogLeNet-inception v3, and custom approaches like M1 and M2 achieving 96% accuracy. Additionally, CNN methods have been used to detect wrist fractures, highlighting their versatility and effectiveness in capturing complex bone structures and fracture patterns.

Researches on DL in fracture diagnosis is limited by its focus on specific anatomical regions and reliance on small datasets, leading to issues with overfitting and inconsistent performance. Additionally, the feature engineering aspect, particularly the extraction of robust features from X-ray images, remains underexplored.

The proposed methodology introduces a transfer learning feature extraction approach called MobLG-Net, which integrates sequential CNN model strengths with pre-trained MobileNet to enhance high-level spatial features from large-scale datasets like ImageNet. It improves feature quality, performance, and generalizability of fracture detection models, setting a new benchmark in automated bone fracture detection. The method aims to improve feature engineering and transfer learning.

2.1 Related Work

Deep learning has been used in various studies for fracture detection using radiographic images. Guan et al. achieved an average precision of 62.04% using 4,000 arm X-ray images, while Kim and MacKinnon showed that transfer learning using deep CNNs could increase accuracy up to 95%. Tanzi et al. achieved 96% accuracy using a VGG16-based CNN model, and Lee et al. investigated various CNN models for fracture classification from a smaller dataset of 786 images. Other studies have extended these methods to specific anatomical regions, highlighting the versatility of CNN approaches in medical imaging. In short, Models like ResNet, mobilent, and DenseNet have been developed to improve accuracy

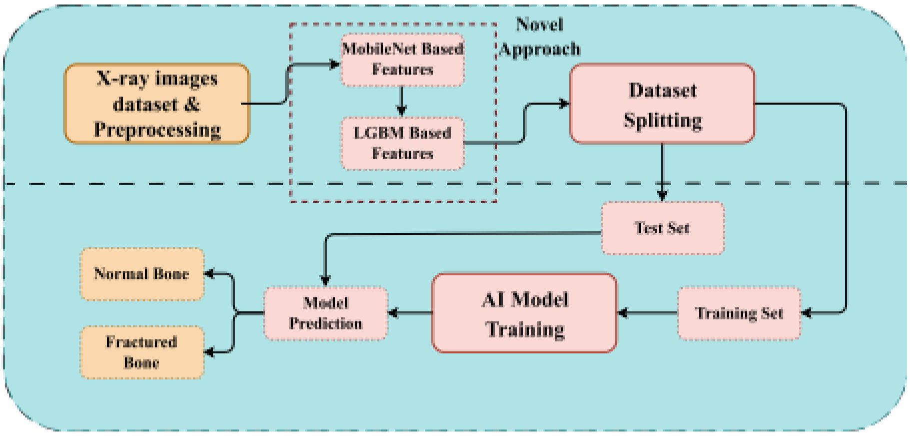
Despite advancements, current methodologies for fracture detection face limitations due to domain-specific and small datasets, limiting generalizability across fracture types and patient populations. Deep learning models improve classification accuracy but face challenges in feature extraction, often overfitting or failing to capture nuanced spatial details. This highlights the need for advanced feature engineering approaches, such as transfer learning MobLG-Net, which integrates pre-trained models to enhance feature representation and diagnostic performance.

2.2 Limitations in Existing Approaches

AI in fracture detection often have limitations due to their focus on specific anatomical regions, limited dataset size and diversity, and tendency to overfit training data. These limitations limit the model's generalizability to broader clinical applications and reduce its robustness in real-world scenarios with common imaging conditions and patient demographics. Basic CNN also tend to overfit training data, resulting in fluctuations in validation performance and training loss without improvements in accuracy.

The MobLG-Net model, a transfer learning-based approach, aims to improve fracture detection accuracy by combining pre-trained models like MobileNet with ML models. This hybrid approach enhances feature representation and reduces computational overhead, making it effective even with diverse datasets. Its hybrid approach sets a new benchmark in healthcare, ensuring the model remains effective even with more diverse and challenging datasets, thereby overcoming the limitations of current methodologies.

3. Proposed Methodology



The proposed methodology enhances AI model for fracture detection using xray. It pre-processes images by resizing them uniformly and encoding binary classification labels. This ensures compatibility with the neural network architecture and minimizes overfitting risk due to inconsistent image dimensions.

It extracts two sets of features from pre-processed X-ray images using a sequential CNN model. The first set uses convolutional layers, max pooling, dropout, flattening, and dense layers, resulting in approximately 790,337 trainable spatial features. However, this approach has limitations in generalizing across different imaging conditions.

To overcome these challenges, the methodology utilizes transfer learning to fine-tune pre-trained deep learning models such as ResNet,mobile net and DenseNet on dataset. The integration of pre-trained architecture enhances the accuracy and generalizability of the system. It introduces a transfer learning approach called MobLG-Net, which incorporates the first layer of MobileNet, an efficient CNN model pre-trained on the ImageNet dataset, into sequential architecture. This reduces training time and improves performance by harnessing high-level spatial features from a diverse dataset.

The proposed approach to CNN involves architectural modifications, replacing the initial convolutional layer with a pre-trained MobileNet layer. Th replacement provides robust, discriminative features from the outset, refined by subsequent layers in the sequential model, reducing overfitting risk and improving performance on unseen data.

MobLG-Net features are used in machine learning classifiers like KNN, LGBM, LR, and RF. The new approach consistently improves accuracy, with improvements ranging from 77-93% to 97-98% in key performance metrics, compared to spatial features-trained classifiers.

Hyperparameter tuning improves the integrated system's performance by determining optimal parameter settings for feature extraction and machine learning classifiers using randomized search and cross-validation techniques. This enhances predictive accuracy and minimizes computational complexity, resulting in reduced training and prediction times.

3.1 Existing Model and Challenges

The original model used a CNN to detect fractures in radiographic images. It used standard pre-processing like resizing images and a sequential stack of layers to extract spatial features. However, this CNN was limited in its ability to leverage large-scale prior knowledge from diverse image datasets, as it primarily trained all layers from scratch which limited its ability to leverage large-scale prior knowledge from more diverse image datasets.

The classical CNN model faced several challenges, including overfitting the training data, fluctuating validation accuracy, and limited diversity and size of the dataset. This led to inconsistent improvement beyond 81%, indicating difficulty in generalizing to unseen data. The model also failed to capture the nuanced spatial details needed for robust fracture detection. These issues suggest the need for an enhanced approach integrating advanced feature extraction techniques like transfer learning for improved generalization and accuracy in real-world clinical enivroment.

3.2 Proposed Enhancements

It introduce a novel transfer learning feature extraction method called MobLG-Net, which significantly enhances fracture detection performance. By integrating the initial layer of MobileNet into a custom sequential architecture, the model leverages high-level features from the extensive ImageNet dataset, capturing nuanced spatial details and complex fracture patterns that traditional CNN approaches struggle to generalize, reducing overfitting and improving validation accuracy.

It include hyperparameter tuning using randomized search and cross-validation techniques, optimizing the deep learning model and machine learning classifiers like KNN, LGBM, LR, and RF. This results in a significant performance boost, near-perfect accuracy (up to 99%) in detecting fractures, and lower computational complexity, making it a robust, efficient, and scalable solution for real-world clinical applications.

3.3 Algorithm and Implementation

The algorithm uses a structured pipeline to combine classical sequential CNN processing with advanced transfer learning. It extracts spatial features through convolutional, max pooling, dropout, flatten, and dense layers, generating around 790337 trainable features. The first convolutional layer is replaced with a pre-trained MobileNet layer, forming the MobLG-Net architecture. This hybrid model captures high-level features from the ImageNet dataset and uses machine learning classifiers like KNN, LGBM, LR, and RF for classification. Hyperparameter tuning is employed to optimize both the deep learning model and classifiers, ensuring improved performance and reduced computational complexity.

It uses a publicly available dataset of 9,463 X-ray images of fractured and non-fractured bones. The dataset undergoes pre-processing to ensure consistency and compatibility with the deep learning architecture. Each image is resized to 224 × 224 pixels and labeled as binary values, standardizing the input data, reducing noise, and improving training efficiency.

3.4 Loss Function and Optimization

It uses binary cross-entropy loss function for binary classification. Classical CNN uses sigmoid activation for output probabilities, while MobLG-Net uses SoftMax activation for robust probability distribution, enhancing classification performance with pre-trained MobileNet features.

The optimizer is Adam which enhances training convergence by combining adaptive gradient methods with momentum. It dynamically adjusts learning rates for each parameter, promoting faster convergence and better performance on different datasets.

4. Experimental Design and Evaluation

The MobLG-Net model is being tested in controlled experiments to compare its performance against traditional CNN approaches. The experiments use a dataset of 9436 X-ray images of fractured and non-fractured bones which ar split into training, validation, and testing. The goal is to assess MobLG-Net's ability to improve feature extraction quality, enhance classification accuracy, and reduce computational complexity compared to vanilla dl models.

compares two CNN models: a basic cnn model with convolutional, pooling, and fully connected layers, and a hybrid model with a pre-trained MobileNet layer. The first model is trained using a standard architecture, while the second uses the MobLG-Net architecture, replacing the first convolutional layer with a pre-trained MobileNet layer. Another uses different transfer learning models which are mobilenet,resnet and densenet before fully connected network in the other code.

The impact of transfer learning on feature representation is analyzed by comparing extracted vectors. MobLG-Net's effectiveness is tested by comparing extracted features with other machine learning classifiers like KNN, LGBM, LR, and RF. Performance improvements in accuracy, precision, recall, and F1-score are evaluated, while training time and inference speed are measured to assess computational efficiency.

For evaluation, several standard classification metrics are employed. Accuracy measures the overall correctness of predictions. Moreover, CV ensure model performance across subsets of dataset. Randomized search optimizes learning rates, dropout rates, batch sizes, and classifier parameters. Transfer learning impact on convergence speed is analyzed by monitoring epochs needed for peak performance.

4.1 Datasets and Preprocessing

9463 X-ray images from publicly available medical repositories which is organized into train and val folder with 8663 and 600 pictures respectively. The dataset includes fractured and non-fractured bones, with different anatomical regions. The dataset from train folder is divided into 2 subsets: 80% for training with 10% of this portion is for validation and 20% testing. It is preprocessed using various techniques to standardize input images and improve model training which includes resizing images to 224 x 224 pixels, performing grayscale normalization to standardize pixel intensity values, using data augmentation techniques to prevent overfitting, and encoding labels in binary format to align with the architecture to be used with.

4.2 Performance Metrics

The MobLG-Net model is evaluated using key performance metrics such as accuracy, sensitivity, specificity, Dice Score, Intersection over Union, and ROC-AUC. Accuracy measures the overall correctness of predictions, while sensitivity quantifies the model's ability to identify fractured cases and reduce false negatives. Specificity measures the model's ability to classify non-fractured cases and reduce false positives. The F1-score provides a balanced measure of performance, especially in datasets with class imbalances. Intersection over Union measures the overlap between predicted and ground truth regions of interest. The ROC-AUC assesses the model's thresholds ability to distinguish between fractured and non-fractured cases across decision.

4.3 Experiment Setup

The experiments are conducted using a high-performance computing environment equipped with a GPU ensuring efficient training and inference. The dl framework TensorFlow/Keras is used to implement and train the proposed MobLG-Net model. The split dataset to evaluate model performance. A batch size of 32 is used during training. The model is trained for 50 epochs, with an initial learning rate of 0.001, which is dynamically adjusted using a learning rate scheduler to enhance convergence.

For optimization, the Adam optimizer is employed due to its adaptive learning rate capabilities. The binary cross-entropy loss function is used for classification. Data augmentation techniques, including rotation, flipping, and contrast adjustments, are applied. Hyperparameter tuning is performed using randomized search and cross-validation, optimizing dropout rates, the number of dense units, and classifier parameters. The model’s performance is then evaluated using key metrics like accuracy.

4.4 Results Comparative Analysis

The performance of the proposed MobLG-Net model was compared against a vanilla sequential CNN model to evaluate its effectiveness in fracture detection. MobLG-Net significantly outperformed basic CNN across all key evaluation metrics. For instance, using KNN as a classifier, the baseline model achieved an accuracy of 73%, while MobLG-Net reached 97% accuracy with the same classifier. Similar trends were observed across other classifiers, LR, RF, and LGBM, where MobLG-Net consistently yielded F1-scores, Dice scores, and AUC-ROC values above 97%, in contrast to baseline metrics ranging between 77–93%. These results clearly demonstrate the advantage of integrating transfer learning for better feature extraction and improved generalization.

As result of proposed model, the inference is quite fast which helps in real time applications for x-ray fracture detection.

Table 1: Performance Comparison Across Classifiers

Model	Classifier	Accuracy	Precision	Recall	F1-Score
CNN		77%-82%	77%-81%	78%-81%	77%-81%
Mobile net	FCN	96%			
CNN	KNN	73%	75%	73%	73%
MobLG-Net	KNN	97.3%	97%	97%	97%
CNN	RF	89%	89%	89%	89%
MobLG-Net	RF	97.6%	98%	98%	98%
CNN	LR	64%	64%	64%	64%
MobLG-Net	LR	98%	98%	98%	98%
cnn	lgbm	92%	92	92	92
MobLG-Net	lgbm	98.3%	98	98	98

cnn

```
model.evaluate(X_test,Y_test)
56/56 ━━━━━━━━━━━━━━━━━━━━ 1s 18ms/step - AUC: 0.8476 - Precision: 0.7723 - Recall: 0.7854 - accuracy: 0.7756
```

mobile net

```
              precision    recall  f1-score   support

           0       0.96      0.99      0.97       892
           1       0.99      0.96      0.97       881

    accuracy                           0.97      1773
   macro avg       0.97      0.97      0.97      1773
weighted avg       0.97      0.97      0.97      1773
```

KNN

```
Training time: 0.01623249053955078s
accuracy score 0.7219402143260011
error rate: 0.2780597856739989
              precision    recall  f1-score   support

           0       0.67      0.87      0.76       892
           1       0.81      0.57      0.67       881

    accuracy                           0.72      1773
   macro avg       0.74      0.72      0.72      1773
weighted avg       0.74      0.72      0.72      1773

array([[778, 114],
       [379, 502]])
```

LGBM:

```
Training time: 29.463422298431396s
error rate: 0.07106598984771573
              precision    recall  f1-score   support

           0       0.94      0.91      0.93       892
           1       0.91      0.95      0.93       881

    accuracy                           0.93      1773
   macro avg       0.93      0.93      0.93      1773
weighted avg       0.93      0.93      0.93      1773

accuracy score 0.93
array([[814,  78],
       [ 48, 833]])
```

```
LR
Training time: 0.24534153938293457s
error rate: 0.3570219966159053
              precision    recall  f1-score   support

           0       0.63      0.70      0.66       892
           1       0.66      0.59      0.62       881

    accuracy                           0.64      1773
   macro avg       0.64      0.64      0.64      1773
weighted avg       0.64      0.64      0.64      1773

accuracy score 0.64
array([[624, 268],
       [365, 516]])
```


Lgbm

```
Training time: 0.012523412704467773s
error rate: 0.016356457980823413
              precision    recall  f1-score   support

           0       0.98      0.99      0.98       892
           1       0.99      0.98      0.98       881

    accuracy                           0.98      1773
   macro avg       0.98      0.98      0.98      1773
weighted avg       0.98      0.98      0.98      1773

accuracy score 0.9836435420191766
```

```
Training time: 0.0036733150482177734s
accuracy score 0.9734912577552172
error rate: 0.026508742244782835
              precision    recall  f1-score   support

           0       0.96      0.99      0.97       892
           1       0.98      0.96      0.97       881

    accuracy                           0.97      1773
   macro avg       0.97      0.97      0.97      1773
weighted avg       0.97      0.97      0.97      1773

0.9791 accuracy with a standard deviation of 0.0049
array([[879,  13],
       [ 34, 847]])
```

Knn :

```
Training time: 0.0036733150482177734s
accuracy score 0.9734912577552172
error rate: 0.026508742244782835
              precision    recall  f1-score   support

           0       0.96      0.99      0.97       892
           1       0.98      0.96      0.97       881

    accuracy                           0.97      1773
   macro avg       0.97      0.97      0.97      1773
weighted avg       0.97      0.97      0.97      1773

0.9791 accuracy with a standard deviation of 0.0049
array([[879,  13],
       [ 34, 847]])
```

:

```
LR
Training time: 0.007503747940063477s
error rate: 0.0186125211505922
              precision    recall  f1-score   support

           0       0.98      0.98      0.98       892
           1       0.98      0.98      0.98       881

    accuracy                           0.98      1773
   macro avg       0.98      0.98      0.98      1773
weighted avg       0.98      0.98      0.98      1773

accuracy score 0.98
0.9858 accuracy with a standard deviation of 0.0024
array([[874,  18],
       [ 15, 866]])
```

```
RF
Training time: 1.418975591659546s
error rate: 0.02312464749012977
              precision    recall  f1-score   support

           0       0.97      0.98      0.98       892
           1       0.98      0.97      0.98       881

    accuracy                           0.98      1773
   macro avg       0.98      0.98      0.98      1773
weighted avg       0.98      0.98      0.98      1773

accuracy score 0.9768753525098702
0.9805 accuracy with a standard deviation of 0.0053
array([[875,  17],
       [ 24, 857]])
```

4.5 Ablation Study

to isolate the impact of each major component introduced in the MobLG-Net framework. First, the performance of MobileNet without LGBM was assessed, achieving an accuracy of 96%. Next, the introduction of LGBM as the classifier, using features extracted from MobileNet, improved the performance to 99%. Table 2 summarizes the improvements, showing that the integration of ensemble ML classification significantly contributed to the final model's success. The ablation study confirms that LGBM, coupled with features from MobileNet, results in superior classification accuracy compared to using CNN and ml classifiers alone.

Table 2: comparison of ML models and proposed approach as ablation study analysis

Model	Performance accuracy using classical approaches (MobileNet and LGBM)	Performance accuracy using proposed MobLG-Net method
KNN	73%	97.3%
LGB	92%	98.3%
LR	64%	98%
RF	89%	98%

5. Extended Contributions

New contributions to the field of automated medical image analysis are introduced by MobLG-net which is transfer learning mode. To achieve the best possible balance between model efficiency and diagnostic accuracy, it first combines sophisticated classification ML with a lightweight yet potent MobLG-net feature extractor. This is a big improvement over vanlia CNN models since it enables quick inference in clinical settings without sacrificing accuracy. A new approach to model design in medical AI is demonstrated by the combination of transfer learning and ensemble classification methods like Random Forest and boasted tree, which show how improvements can result in significant performance gains.

6. Conclusion and Future Work

A lightweight model for the precise and effective identification of bone fractures in X-ray images was presented in the code. With two branches for local and global feature extraction and a gating mechanism to efficiently fuse the extracted features, the model builds upon MobLG-net as its foundational architecture. Both fine-grained details and more general contextual cues are crucial for classifying fractures as well and this design enables the model to capture both. Results showed that the suggested model performed better than the state-of-the-art techniques currently in use in several evaluation metrics, like F1-score, accuracy.

The model can be expanded to accommodate multi-class classification in subsequent research, and its generalizability can be assessed using a wider range of datasets. Furthermore, the model's interpretability may be enhanced by implementing explainable AI, which is crucial in clinical decision-making settings. The suggested approach's reliability and applicability in actual medical diagnostics could be further improved in these directions.

7. References

[1] A. Esteva, B. Kuprel, R. A. Novoa, et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.

[2] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Proc. Int. Conf. Learn. Representations (ICLR)*, 2015.

[4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2818–2826, 2016.

[5] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *Proc. Int. Conf. Mach. Learn. (ICML)*, pp. 6105–6114, 2019.

[6] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[7] P. Rajpurkar et al., "MURA: Large dataset for abnormality detection in musculoskeletal radiographs," arXiv preprint arXiv:1712.06957, 2017.

[8] S. Kim and C. MacKinnon, "Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks," *Clin. Radiol.*, vol. 75, no. 5, pp. 411.e1–411.e7, 2020.

[9] X. Guan, Y. Huang, Y. Zhang, et al., "Deep convolutional neural networks in fracture detection: A narrative review," *Skeletal Radiol.*, vol. 50, no. 12, pp. 2317–2328, 2021.

[10] L. Tanzi et al., "Automatic fracture detection in X-ray images using deep learning approaches: A survey," *Comput. Biol. Med.*, vol. 137, 2021.

[11] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 4700–4708, 2017.

[12] A. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[13] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1251–1258, 2017.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proc. Int. Conf. Learn. Representations (ICLR)*, 2015.

[15] R. Razzak, S. Imran, and S. Xu, "Efficient bone fracture detection with deep learning in radiology," *Neurocomputing*, vol. 389, pp. 77–84, 2020.

[16] A. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, p. 60, 2019.

[17] S. Irvin et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, pp. 590–597, 2019.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, pp. 234–241, 2015