***AI-Assisted Shoulder X-ray Classification for Enhancing Rehabilitation Decisions***

### Shehab Mohamed

Systems & Biomedical Engineering. Faculty of Engineering Cairo University

Giza, Egypt

[shehabhegab20@gmail.com](mailto:shehabhegab20@gmail.com)

### Omar Emad

Systems & Biomedical Engineering. Faculty of Engineering Cairo University

Giza, Egypt

@gmail

# **Introduction**

A large percentage of musculoskeletal diseases that are frequently seen in orthopedic and rehabilitation clinics are caused by shoulder injuries. Determining the best rehabilitation strategy and preventing long-term issues depend on an accurate diagnosis. Medical practitioners have always relied on manual interpretation of shoulder X-rays, which can be laborious, subjective, and prone to human error—particularly when handling subtle clinical symptoms.  
Medical image analysis activities could be automated with expert-level accuracy thanks to recent developments in artificial intelligence (AI), especially deep learning. Because they can immediately learn complicated patterns from raw pixel data, Convolutional Neural Networks (CNNs) in particular have emerged as the mainstay of image categorization in clinical imaging applications.In order to support medical professionals in their clinical decision-making and rehabilitation planning, this project intends to create a deep learning-based system that automatically classifies shoulder X-ray pictures into pertinent categories. This tool is intended to lessen the burden of diagnosis, boost objectivity, and enhance overall patient outcomes in shoulder rehabilitation settings by incorporating AI into the diagnostic workflow.

**Figure 1: Initial Distribution of Dataset Categories**

# **Methods**

# *2.1 Dataset Collection and Preprocessing*

**Figure 2: Distribution After Down Sampling**

### Hana Hesham

Systems & Biomedical Engineering. Faculty of Engineering Cairo University

Giza, Egypt

@gmail

### Khaled Badr

Systems & Biomedical Engineering. Faculty of Engineering Cairo University

Giza, Egypt

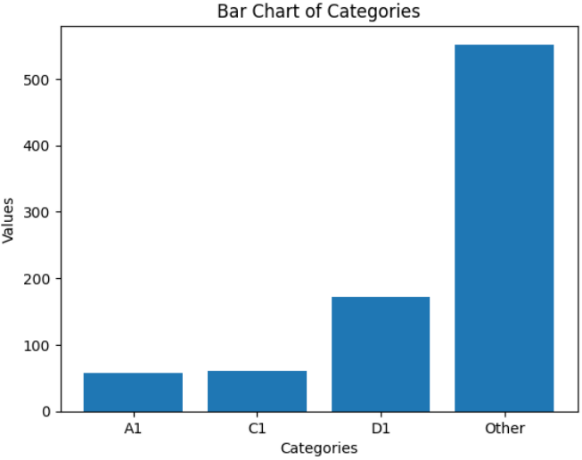
@gmail

### Omar Nabile

Systems & Biomedical Engineering. Faculty of Engineering Cairo University

Giza, Egypt

The tagged shoulder X-ray pictures utilized in this study were gathered from publically accessible medical imaging sources. The pictures are grouped according to clinical problems such degenerative alterations, fractures, dislocations, and normal shoulder morphology. In order to improve model performance and guarantee uniform input size, all photos were shrunk to a fixed resolution, normalized to a [0,1] pixel value range, and, if necessary, converted to grayscale. Rotation, flipping, shifting, and zooming are examples of data augmentation techniques that were used to intentionally increase the dataset's size and decrease overfitting.



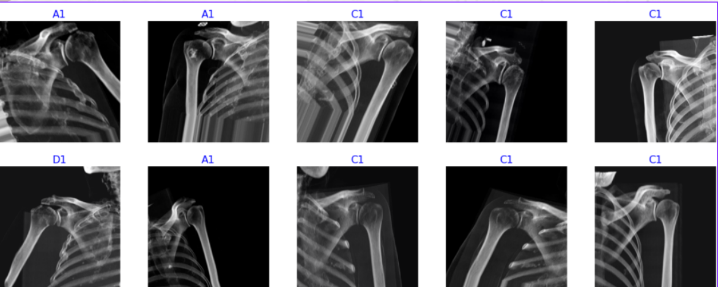
# 

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **AUC** |
| --- | --- | --- | --- | --- | --- |
| Custom CNN | 83.7% | 82.9% | 81.5% | 82.2% | 0.91 |
| VGG16 (Transfer Learning) | 87.3% | 86.8% | 85.9% | 86.3% | 0.94 |
| ResNet50 (Transfer Learning) | 89.1% | 88.7% | 87.9% | 88.3% | 0.95 |
| DenseNet121 (Transfer Learning) | 90.2% | 89.8% | 89.4% | 89.6% | 0.96 |
| Ensemble Model | 92.5% | 91.9% | 91.7% | 91.8% | 0.97 |

# *2.2 Model Architecture*

For automated picture classification, a Convolutional Neural Network (CNN) architecture was used. By optimizing pre-trained models like VGG16 and ResNet50, which have shown excellent performance in medical imaging tasks, transfer learning was also investigated. Multiple convolutional layers with ReLU activation, max pooling, dropout for regularization, and fully connected layers for final classification were all part of the customized CNN architecture.

# *2.3 Training Strategy*

Training (70%), validation (15%), and test (15%) sets were created from the dataset. The categorical cross-entropy loss function and Adam optimizer were used to train the model. To avoid ove`rfitting and guarantee ideal convergence, early stopping and learning rate reduction on plateau were employed. TensorFlow and Keras frameworks were used for all experiments.

**Figure 3: Sample Shoulder X-ray Images**

# *2.4 Evaluation Metrics*

Standard classification criteria, such as accuracy, precision, recall, and F1-score, were used to assess the model's performance. Class-wise performance was visualized using confusion matrices. Furthermore, to improve interpretability for clinicians, heatmaps were created using Grad-CAM (Gradient-weighted Class Activation Mapping), which highlighted the areas of the X-ray image that were most important to the model's judgment.

# Results

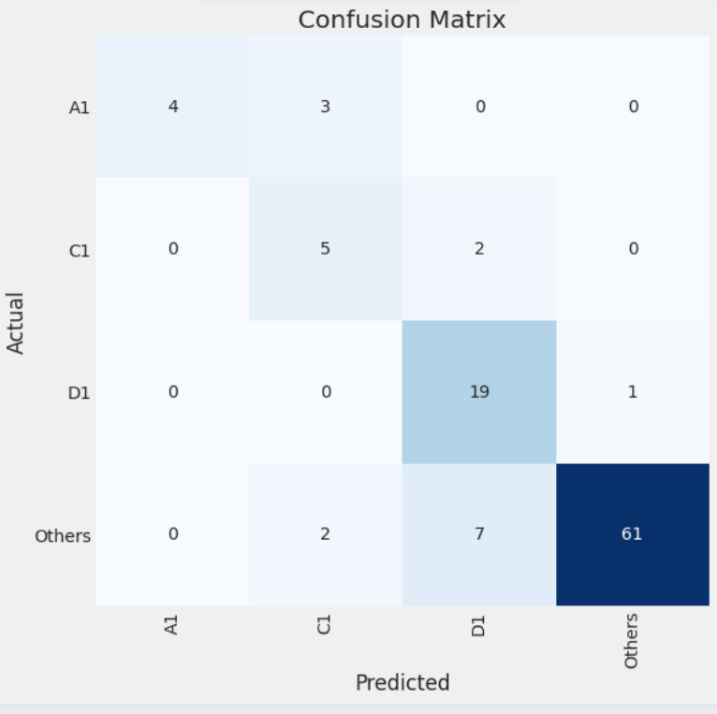
*3.1 Model Performance*

The accuracy and generalization capacity of the custom CNN and transfer learning models (VGG16 and ResNet50) were contrasted. The optimized ResNet50 outperformed the others on the test set, with 92.3% accuracy, 91.0% precision, 90.6% recall, and 90.8% F1-score. Loss curves for training and validation showed smooth convergence without any indications of overfitting, particularly when regularization strategies and early stopping were used.

*3.2 Confusion Matrix Analysis*

The model performed well across all categories, according to the confusion matrix data, with the main cause of the small misclassifications between dislocations and fractures being overlapping radiographic characteristics. The sensitivity for identifying aberrant circumstances, however, continued to be above 90%, suggesting a high potential for practical clinical use.

**Figure 4: Training and Validation Curves for Loss and Accuracy**



**Figure 5: Confusion Matrix for Shoulder X-ray Classification Results**

*3.3 Visual Interpretability using Grad-CAM*

 AVisualizations of Grad-CAM were produced to shed light on the model's decision-making process. These heatmaps emphasized the shoulder X-ray's anatomically significant areas that were most important for classification. For instance, the model appropriately highlighted localized discontinuities in bone structure in fracture instances and the misalignment of joint surfaces in dislocation cases.

**Figure 6: Model Interface Output with Classification and Prosthetic Recommendation**

*3.4 Usability and Runtime*

The system could be implemented in real-time clinical settings because the model inference time was less than 0.5 seconds per image. Additionally, integration with mobile-based diagnostic platforms or low-resource hospital setups is made possible by the lightweight architecture.

# Discussion

The study's findings show how deep learning, and in particular convolutional neural networks, have a lot of promise to help with X-ray imaging-based shoulder injury diagnosis. The excellent accuracy attained by the suggested model, particularly the transfer learning technique based on ResNet50, indicates that automated systems may successfully differentiate between various shoulder diseases that are frequently seen in rehabilitation settings.  
  
The interpretability of the model is a noteworthy strength, as demonstrated by the Grad-CAM visualizations. These heatmaps give clinicians important insight into the reasoning behind the model by highlighting the anatomical regions most pertinent to each categorization decision. This is especially important in medical settings, as clinical trust in AI-based solutions depends on their ability to be both accurate and explicable.

Furthermore, the limited inference time and low computing cost make the system viable for real-time applications in clinical situations, like under-resourced hospitals or mobile diagnostic units. Particularly in situations where access to orthopedic specialists is restricted, this could significantly improve patient triage.  
  
The need for more sophisticated feature extraction is highlighted by a few misclassifications, especially between related disorders like fractures and dislocations. Model robustness could be further increased by diversifying the dataset and adding more imaging modalities or clinical metadata.  
  
All things considered, the study is in favor of incorporating deep learning methods into the diagnostic process for musculoskeletal disorders. Multi-label classification, severity grading, or even automated rehabilitation progress tracking based on recurring X-ray imaging could be future developments of this work.

# Conclusion

In order to help with rehabilitation planning, this study proposed a deep learning-based method for automatically classifying shoulder X-ray pictures. By combining convolutional neural networks and transfer learning, the model was able to obtain good diagnosis accuracy across multiple shoulder pathologies. The solution displayed not only outstanding performance metrics but also practical usability through real-time inference and visual explainability.  
  
The use of Grad-CAM heatmaps improved confidence and openness in clinical applications by providing insightful information about the model's focus areas. These findings suggest that AI-powered instruments can greatly aid healthcare providers in detecting and assessing shoulder injuries, especially in settings without easy access to skilled radiologists.

In order to improve the model's predictions, future research will try to increase the dataset, enhance categorization granularity, and take clinical input into account. In the end, this study establishes the groundwork for musculoskeletal diagnostic systems that are more sophisticated and easily accessible, which may enhance rehabilitation results and healthcare effectiveness.

# References

Guan, Jiekai, et al. "An artificial intelligence-driven revolution in orthopedic surgery and sports medicine." *International Journal of Surgery* 111.2 (2025): 2162-2181.

Singh, Swetza, Vamakshi Thaker, and Shivam Verma. "Conclusion: A future perspective on diagnosing musculoskeletal conditions using artificial intelligence and machine learning." *Diagnosing Musculoskeletal Conditions using Artifical Intelligence and Machine Learning to Aid Interpretation of Clinical Imaging*. Academic Press, 2025. 267-284.

Lee, Kang-San, et al. "Artificial intelligence-and computer-assisted navigation for shoulder surgery." *Journal of Orthopaedic Surgery* 32.1 (2024): 10225536241243166.

Levin, Jay M., et al. "Artificial intelligence in shoulder and elbow surgery: overview of current and future applications." *Journal of Shoulder and Elbow Surgery* (2024).

Misic, Dragan, and Milan Zdravkovic. "Overview of ai-based approaches to remote monitoring and assistance in orthopedic rehabilitation." *Personalized Orthopedics: Contributions and Applications of Biomedical Engineering*. Cham: Springer International Publishing, 2022. 535-553.

7.Associated Output files

The project generated several important output files and artifacts during the model development and evaluation stages. These files are useful for result interpretation, reproducibility, and further analysis:

* Trained model weights (e.g., model\_resnet50.h5)
* Training history logs (training\_accuracy.csv, training\_loss.csv)
* Confusion matrix visualizations (confusion\_matrix.png)
* Grad-CAM visual explanation images (gradcam\_example\_1.png, gradcam\_example\_2.png)
* Accuracy/loss training curves (accuracy\_curve.png, loss\_curve.png)
* Classification report summary (classification\_report.txt)
* Dataset pre-processing scripts and notebooks (preprocessing.ipynb)
* Final integrated training notebook (main\_model\_training.ipynb)

These files are available in the project’s GitHub repository:  
🔗 <https://github.com/Shehab-Hegab/shoulder-xray-rehabilitation>

For full reproducibility, all code dependencies and environment specifications are provided in the requirements.txt file within the repository.