Toward a Unified Public X-ray Dataset Integrating Multiple Databases to Advance Complex Fracture Analysis

Fatma Atitallah[1], Assem Thabet[1], Johannes C. Ayena[2,3], Neila Mezghani[2,3]

[1]Research Laboratory MACS, University of Gabès, Gabès, Tunisia
e-mail: fatma.atitallah@enig.u-gabes.tn, assem.thabet@issig.rnu.tn

[2]Applied Artificial Intelligence Institute (I2A), TELUQ University, Montreal, Quebec, Canada
[3]Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montreal, Quebec, Canada
e-mail: johannes.ayena@teluq.ca, neila.mezghani@teluq.ca

***Abstract*—Detecting and segmenting complex upper-limb fractures in X-ray imaging remains a persistent challenge due to subtle visual patterns, anatomical overlap, and heterogeneous acquisition conditions. While deep learning continues to advance fracture analysis, its progress is limited by fragmented public datasets that differ in format, annotation standards, clinical focus, and image quality. This paper addresses this limitation by unifying three publicly accessible X-ray datasets (FracAtlas, Bone Fracture Detection, and GRAZPEDWRI-DX) into a single, standardized, high-quality resource dedicated to upper-limb fracture research. The proposed integration pipeline includes rigorous dataset selection, annotation harmonization, preprocessing, normalization, and multi-stage quality control to ensure consistency and clinical reliability. The resulting dataset provides extensive anatomical diversity, a wide spectrum of fracture types, and verified segmentation masks suitable for training and benchmarking deep learning models. Although class imbalance and uneven fracture representation persist, the unified dataset establishes a valuable foundation for developing robust, generalizable, and clinically meaningful Artificial Intelligence (AI) systems for fracture detection, localization, and segmentation. Future extensions will focus on balancing fracture categories and evaluating state-of-the-art architectures on the proposed dataset.**

***Keywords*-Upper-limb fractures; X-ray imaging; Dataset harmonization; Medical image preprocessing; Fracture detection; Fracture segmentation; Artificial Intelligence; Dataset integration.**

I. Introduction

Precisely detecting and segmenting complex bone fractures in X-ray images remains an essential yet highly challenging task in medical imaging, due to the subtle nature of fractures, the frequent overlap of anatomical structures, and the considerable variation in image quality [1]. Upper-limb fractures, including those of the fingers, hands, wrists, forearms, and elbows, can be particularly challenging due to the complex anatomy of these areas, the small size of some bones, and the wide variety of fracture types, such as displaced, comminuted, or hairline fractures [2]. Errors or delays in diagnosing such fractures can lead to prolonged recovery periods, reduced functional mobility, and increased healthcare costs [3]. X-ray imaging continues to be the primary method for evaluating fractures due to its speed and accessibility [4][5]. Despite these benefits, interpreting X-rays accurately remains a research challenge because it depends largely on the clinician's expertise [1]. Studies have shown significant variability between observers, especially when evaluating subtle or rare fractures, highlighting the need for objective and standardized tools to support clinical decision-making [4]. Artificial Intelligence (AI) and deep learning have shown considerable promise in medical imaging, helping clinicians more reliably detect, locate, and segment fractures [6][7]. However, building effective AI models depends on access to large, diverse, and high-quality datasets [8][9]. At present, existing X-ray datasets are often fragmented, vary in format, and lack consistent annotations, which limits their practical use for model development [10]. This study proposes to address these limitations by integrating, standardizing, and harmonizing multiple publicly available X-ray datasets into a unified resource. In doing so, we create a comprehensive, clinically meaningful, and fully standardized X-ray dataset dedicated to upper-limb fracture analysis.

The novelty of this work lies in the systematic integration of three major publicly available datasets FracAtlas [10][16], Bone Fracture Detection (Kaggle) [17], and GRAZPEDWRI-DX [18] into a single, coherent resource. Publicly available datasets that are fully annotated for fracture segmentation are rare, and identifying datasets that simultaneously provide X-ray imaging, upper-limb coverage, pixel-level segmentation annotations, and comprehensive metadata remains a significant challenge. Among the publicly accessible resources, only these three datasets currently meet our selection criteria. While other fracture related datasets do exist, they are often limited to Computed Tomography (CT) imaging modalities or do not cover the upper limb, such as RSNA [11] and 5K+ CT Images on Fractured Limbs [12], or are primarily designed for classification tasks, such as the MURA dataset [0], rather than segmentation, making them unsuitable for the objectives of this work.

This scarcity highlights the uniqueness of such datasets and directly motivates the core objective of this study: the construction of a unified, balanced, and high-quality data distribution from heterogeneous public sources. To achieve this objective, the proposed approach harmonizes annotation formats, imaging resolutions, metadata structures, and preprocessing pipelines, while implementing a multi-stage quality control procedure to guarantee data consistency and reliability.

By creating a reliable and comprehensive dataset, this work establishes a strong foundation for the development of AI systems capable of supporting clinicians in accurate and efficient fracture diagnosis, improving patient care, and advancing medical imaging research for the detection and segmentation of complex bone fractures.

The paper is structured as follows: Section II details the methodology, introduces the datasets with their key strengths and limitations, and describes their preparation before fusion. Fusion, Standardization, and Normalization of Datasets covers the merging and preprocessing steps. Dataset Quality Assurance presents post-fusion evaluation, including automated and visual checks. Section III reports and discusses the results, key observations, and trends. Finally, Section IV concludes and outlines future research directions.

II. METHODS

Selecting an appropriate imaging modality is essential for analyzing complex or atypical bone fractures. Public datasets enable the integration of heterogeneous sources into a unified database for automated fracture detection and segmentation [13], ensuring clinical relevance and compatibility with machine learning workflows [14][15]. This study used three publicly accessible datasets, FracAtlas, the Kaggle Bone Fracture Detection dataset, and GRAZPEDWRI-DX (Table I), chosen for their availability, rich annotations, and strong focus on upper-limb fractures. All datasets underwent fusion, standardization, normalization, and quality control to ensure consistency.

The datasets FracAtlas, Bone Fracture Detection, and GRAZPEDWRI-DX were selected primarily based on their public accessibility and their suitability for fracture detection and segmentation tasks in X-ray imaging. The selection focused on datasets covering the upper-limb, which remains underrepresented in publicly available fracture datasets.

To develop robust and generalizable AI models, datasets representing diverse populations and clinical contexts were prioritized. FracAtlas, Kaggle Bone Fracture Detection, and GRAZPEDWRI-DX offer complementary anatomical coverage, detailed annotations, and demographic variability. Together, they form a cohesive resource for training and evaluating fracture detection and segmentation algorithms. Table I summarizes their characteristics, including population type, anatomical coverage, size, annotation formats, metadata, and geographic origin.

A. Datasets Preparation

Prior to dataset integration, the three selected datasets were manually examined. The FracAtlas dataset comprises only 717 fracture images (mixed upper and lower limp), while the other two datasets contain substantially more (as illustrated in Table I). To mitigate this imbalance, data augmentation was applied to the FracAtlas images after selecting only upper-limb samples tripling their number.

Additionally, 3,000 images were randomly selected from each of the GRAZPEDWRI-DX and Bone Fracture Detection datasets before preprocessing to reduce their size. This procedure ensured a more balanced selection across the three datasets prior to fusion.

1) **Data Preparation for FracAtlas:**
The FracAtlas dataset contains X-ray images with pixel-level segmentation masks for precise fracture detection. The dataset is structured into *Images* (Fractured and Non-fractured), *Annotations* (COCO JSON utilized in this study), and *Metadata* recorded in a CSV file. Let N_{init} denote the total number of annotated fractured images before filtering. We first filter the dataset to retain only upper-limb fractures (hand and shoulder) and exclude lower-limb regions (leg and hip). Formally, the filtering condition can be expressed as:

$$\text{UpperLimbFra} = \left\{ i \;\middle|\; (\text{hand}_i = 1 \;\vee\; \text{shoulder}_i = 1) \right\} \wedge (\text{leg}_i = 0 \;\wedge\; \text{hip}_i = 0) \wedge (\text{fractured}_i = 1) \tag{1}$$

where:

- i indexes each image in the dataset,
- $\vee$ (logical OR) indicates that at least one of the conditions is true,
- $\wedge$ (logical AND) indicates that all conditions must be true simultaneously.

This defines *UpperLimbFra* as the set of images showing fractures in the upper limbs (hand or shoulder) while excluding any fractures in the legs or hips.
Each image I_i is associated with a binary mask M_i, where fracture pixels are assigned a value of 255 and background pixels are set to 0 [19], as formally defined by:

$$M_i(x,y) = \begin{cases} 255, & \text{if the pixel } (x,y) \\ & \text{belongs to a fracture,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Applying the filter defined in (1) reduced the dataset to N_{fil} (number of filtered upper-limb fracture images). This filtered set restricts that only relevant images containing fractures in the upper extremities (hand or shoulder) and excluding lower-limb involvement are retained for subsequent analysis and augmentation.
Moreover, images without valid masks ($\sum M_i = 0$) were removed. Data augmentation, as described in [20], was applied to each filtered image. Let A denote the number of augmented images per original sample. In this study, the augmentation factor was fixed to $A = 3$, a choice that provides additional variability without disproportionately amplifying any particular subset of the dataset. This specific choice helps maintain a comparable number of images in the filtered FracAtlas subset relative to the other two datasets, preventing any single dataset from dominating the combined database and supporting a balanced representation for training. The augmentation techniques applied include:

TABLE I. COMPARATIVE SUMMARY OF THE THREE PUBLICLY AVAILABLE FRACTURE DATASETS, INCLUDING DATASET ORIGIN.

Parameter	FracAtlas	Kaggle Bone Fracture Detection	GRAZPEDWRI-DX
Population	Mixed adults & children	Unspecified	Pediatric only (0–19 yrs)
Dataset origin / Country	Bangladesh (3 hospitals)	N/A	Austria (University Hospital Graz)
Body regions	Upper & lower limbs	Mainly upper limbs	Wrist only
Dataset size	4,083 images; 717 with fractures	8296 images	20,327 images; 6,091 patients
Annotations	Classification labels, masks, bounding boxes	Bounding boxes, some masks	Masks, polygons, bounding boxes
Metadata	Available	None	Rich; includes clinical descriptors
Strengths	• Clinically reviewed • Subtle fractures included • Multi-hospital collection	• Wide upper-limb variety • Useful for detection models (YOLO)	• Large-scale pediatric dataset • Detailed annotations • Masks, polygons, bounding boxes
Limitations	• Underrepresented regions • Class imbalance	• Missing metadata • Unclear labeling • Integration harder	• Single-center dataset • Uneven class distribution

Note. yrs: years; N/A: Not Available

- Horizontal and vertical flips applied with probabilities $p_{\text{flipH}} = 0.5$ and $p_{\text{flipV}} = 0.3$.
- Random rotation by an angle $\theta \in \{-15°, -10°, -5°, 5°, 10°, 15°\}$.
- Brightness and contrast modification following $I_{\text{aug}} = \alpha I_{\text{orig}} + \beta$, with brightness factor $\alpha \in [0.8, 1.2]$ and contrast shift $\beta \in [-20, 20]$.
- Additive Gaussian noise: $I_{\text{aug}} = I_{\text{aug}} + \mathcal{N}(0, \sigma^2)$ with noise level $\sigma = 10$.

Here, I_{orig} denotes the original image, and I_{aug} its augmented version. The parameters α and β control brightness and contrast, respectively. The variable θ represents the randomly selected rotation angle, while p_{flipH} and p_{flipV} denote the probabilities of horizontal and vertical flipping. Finally, $\mathcal{N}(0, \sigma^2)$ refers to zero-mean Gaussian noise with variance σ^2.

2) Data Preparation for Bone Fracture Detection from Kaggle:

The Kaggle Bone Fracture dataset, provided in YOLOv8 format, was preprocessed to support both fracture detection and segmentation tasks. This dataset contains seven classes: 'elbow positive', 'fingers positive', 'forearm fracture', 'humerus fracture', 'humerus', 'shoulder fracture', and 'wrist positive', as reported in [17]. The preprocessing pipeline included generating binary masks, computing image-level statistics, and creating metadata for each image.

- **Binary Masks for Segmentation**

  For each image I_i and its corresponding annotations L_i, a binary mask M_i was generated via (2).

  Bounding boxes in YOLO format, with normalized coordinates (x_c, y_c, w, h), were converted to pixel coordinates as follows:

$$\begin{aligned} x_{\min} &= \max(0, x_c \cdot W - \frac{w \cdot W}{2}) \\ x_{\max} &= \min(W - 1, x_c \cdot W + \frac{w \cdot W}{2}) \\ y_{\min} &= \max(0, y_c \cdot H - \frac{h \cdot H}{2}) \\ y_{\max} &= \min(H - 1, y_c \cdot H + \frac{h \cdot H}{2}) \end{aligned} \tag{3}$$

  where W and H denote the image width and height. When available, polygon annotations were rasterized into the mask for more precise fracture localization as represented by:

$$M_i(x, y) = \begin{cases} 255, & (x, y) \in \text{polygon interior} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

  All masks were then binarized as:

$$M_i(x, y) = \begin{cases} 255, & M_i(x, y) > 127 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

  Images without valid annotations or with empty masks were excluded.

- **Metadata Creation**

  For each retained image, a metadata CSV file was generated containing:

  – image name I_i,
  – fracture class for each annotation c_j,
  – annotation coordinates (x_j, y_j),
  – approximate fracture area formally defined as:

$$A_j = (x_{\max} - x_{\min}) \times (y_{\max} - y_{\min}) \tag{6}$$

  Image-level statistics were computed:

  – Number of fractures: $n_i = \text{count}(c_j)$
  – Mean coordinates: $\bar{x}_i = \frac{1}{n_i} \sum_j x_j, \quad \bar{y}_i = \frac{1}{n_i} \sum_j y_j$
  – Mean area: $\bar{A}_i = \frac{1}{n_i} \sum_j A_j$

– Severity category:

$$\text{severity}(I_i) = \begin{cases} \text{mild}, & n_i = 1 \\ \text{multiple}, & n_i > 1 \end{cases} \quad (7)$$

- **Visualization and Statistics**
  The percentage distribution of fracture types was computed as:

$$p_k = \frac{\text{count}(c_j = k)}{N_{\text{final}}} \times 100 \quad (8)$$

where k denotes a possible fracture class value, and N_{final} is the final number of image–mask pairs after preprocessing.

3) **Data Preparation for GRAZPEDWRI-DX:** For preprocessing, a subset of images N_{init} was selected from the GRAZPEDWRI-DX dataset part 1 to maintain balance when merging with other datasets. A fixed random seed allowed reproducible results [4][18].
   The preprocessing pipeline included the following steps:
   - Defining input paths for raw images and Pascal VOC annotations, and creating output folders for preprocessed images and masks.
   - Parsing each annotation to extract polygons or bounding boxes, which were used to generate precise binary masks (0 = background, 255 = fracture).
   - Clamping and resizing masks when necessary to match the image dimensions.
   - Saving the masks alongside images with harmonized filenames.
   - Recording metadata, including filenames, mask names, and classes, in a CSV file.

B. Fusion, Standardization, and Normalization of Datasets

In order to build a comprehensive and consistent dataset, multiple fracture datasets were fused into a single collection while preserving demographic labels (Adult, Child, or Mixed—Adult and Child). Each image I_i and its corresponding mask M_i were resized (resampled) to a standard resolution of 512×512 pixels [21], resulting in I_i^{res} and M_i^{res}, the resized image and mask, respectively. This process is formally defined by:

$$\begin{aligned} I_i^{\text{res}} &= \text{Res}(I_i, 512, 512, \text{interp} = \text{cubic}), \\ M_i^{\text{res}} &= \text{Res}(M_i, 512, 512, \text{interp} = \text{nearest}) \end{aligned} \quad (9)$$

Cubic interpolation preserves image details [22], while nearest-neighbor interpolation maintains mask accuracy [23]. X-ray images often exhibit uneven brightness or shading. To correct this, the N4ITK Bias Field Correction was applied to the luminance channel of each image as formally defined as:

$$\hat{I}_i = I_i^{\text{res}} \cdot B_i^{-1}, \quad B_i = \text{BiasField}(I_i^{\text{res}}) \quad (10)$$

where B_i is the estimated bias field. After bias correction, intensity normalization was performed to scale the pixel values to a consistent range suitable for machine learning, as represented by:

$$I_i^{\text{norm}} = \frac{\hat{I}_i - \mu_{\hat{I}_i}}{\sigma_{\hat{I}_i}} \cdot 255 \quad (11)$$

Here, $\mu_{\hat{I}_i}$ and $\sigma_{\hat{I}_i}$ are the mean and standard deviation of the bias-corrected image $\hat{I}_i$. This normalization ensures consistent brightness and contrast across all images. Preprocessing was performed in batches with parallel processing for efficiency. Metadata, such as dataset source, demographic category, and mask information, were saved alongside the images.

The resulting fused $\mathcal{D}_{\text{fused}}$ and normalized dataset I_i^{norm} is formally defined as:

$$\mathcal{D}_{\text{fused}} = \{(I_i^{\text{norm}}, M_i^{\text{bin}}, \text{metadata}_i)\}_{i=1}^{N_{\text{total}}} \quad (12)$$

where N_{total} denotes the total number of image–mask pairs in the fused dataset, computed as the sum of all filtered and augmented images across the three constituent datasets.

C. Dataset Quality Assurance

After merging the datasets, we ensured consistency and reliability through automated verification, visual inspection, and quantitative analysis. We implemented a standardized integration pipeline that included annotation harmonization, intensity normalization, and multi-stage quality control. These steps were specifically designed to minimize potential data distortions during the merging process and to preserve all clinically relevant fracture features, despite variations in imaging protocols, anatomical coverage, and acquisition quality across the original datasets. This approach guarantees that the resulting unified dataset maintains both anatomical fidelity and suitability for training robust and generalizable deep learning models. Automated checks confirmed that each image had a readable, correctly sized, and non-empty mask, with any anomalies logged for review.

Additionally, a random subset of images was manually inspected by overlaying the fracture regions on the corresponding X-rays to verify accurate mask delineation. To quantify fracture coverage, we calculated the proportion of each image occupied by fractures using (13):

$$\text{Fracture\_\%}_i = \frac{\sum_{x,y} \mathbb{1}[M_i(x,y) > 127]}{\text{width}_i \times \text{height}_i} \times 100, \quad (13)$$

Here, M_i denotes the standardized binary mask of image I_i from its original dataset (before integration into the unified database), and $\mathbb{1}$ is the indicator function counting fracture-labeled pixels. Using these values, we computed standard descriptive statistics for each dataset, including minimum, maximum, mean, and median fracture coverage. To assess fracture complexity, we analyzed the morphology of each connected component in the mask. For a component C_j in M_i with area A_j, width w_j, height h_j, and perimeter P_j, its complexity was computed using (14):

TABLE II. UNIFIED SUMMARY OF THE THREE DATASETS: SIZES, FILTERING, AUGMENTATION, SURFACE STATISTICS, AND COMPLEXITY.

Dataset	N_{init}	**Filtering & Resulting Outputs**			**Surface Statistics (%)**				**Complexity**
		N_{fil}	A	N_{final}	**Min**	**Max**	**Mean**	**Median**	**Mean / Max**
FracAtlas	717	439	3	1,756	0.007	2.241	0.352	0.312	11.48 / 28.66
Kaggle	3,000	1,491	–	1,491	0.123	18.583	2.256	1.651	11.20 / 24.97
GRAZPEDWRI-DX	3,000	2,017	–	2,017	0.320	14.986	2.622	2.288	11.56 / 24.63

Notes: N_{init}: initial number of X-ray images selected in each dataset; N_{fil}: number of valid image–mask pairs remaining after filtering; A: number of augmentation operations applied (only FracAtlas was augmented with $A = 3$); N_{final}: final number of image–mask pairs after filtering and augmentation; *Min*, *Max*, *Mean*, *Median* (%): proportion of fracture pixels; *Mean / Max of Complexity*: average and maximum fracture complexity based on morphological descriptors (connected components, aspect ratio, compactness).

$$\text{Complexity}(C_j) = 0.5 \cdot \frac{\max(w_j, h_j)}{\min(w_j, h_j)} + 0.5 \cdot \frac{P_j^2}{A_j}. \quad (14)$$

The overall complexity of image I_i was then defined as the mean complexity across its n_i components through (15):

$$\text{Complexity}(I_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} \text{Complexity}(C_j). \quad (15)$$

Finally, for each dataset, we computed the mean and maximum complexity across all images retained after preprocessing, using (16):

$$\begin{aligned} \text{Mean Complexity} &= \frac{1}{N_{\text{final}}} \sum_{i=1}^{N_{\text{final}}} \text{Complexity}(I_i), \\ \text{Max Complexity} &= \max_{i \le N_{\text{final}}} \text{Complexity}(I_i). \end{aligned} \quad (16)$$

These metrics provide a comprehensive description of fracture morphology by capturing size and shape irregularities, offering a clear overview of dataset properties for downstream analysis and model development.

III. RESULTS AND DISCUSSION

The filtering procedure defined in (1) reduced the initial FracAtlas dataset from $N_{\text{init}} = 717$ to $N_{\text{fil}} = 439$ upper-limb fracture images. To preserve class balance and dataset diversity, each retained image I_i and its corresponding mask M_i underwent three augmentation operations ($A = 3$), resulting in a final total of $N_{\text{final}} = 1{,}756$ aligned image-mask pairs, as summarized in Table II.

Figure 1 shows the distribution of fracture types after filtering and augmentation, highlighting the predominance of hand fractures and confirming the effective exclusion of leg and hip cases. This ensures the dataset remains representative of upper-limb fractures and suitable for subsequent training and evaluation.

Similarly, the Kaggle Bone Fracture dataset underwent the same preprocessing. Binary masks were generated via (5) and only valid image–mask pairs were retained. Out of 3,000 initially selected images, 1,491 pairs remained. Metadata were then created, and fracture type percentages calculated using (8) to support statistical analysis and task integration. Figure 2 illustrates class imbalance, particularly for wrist-positive and humerus cases.

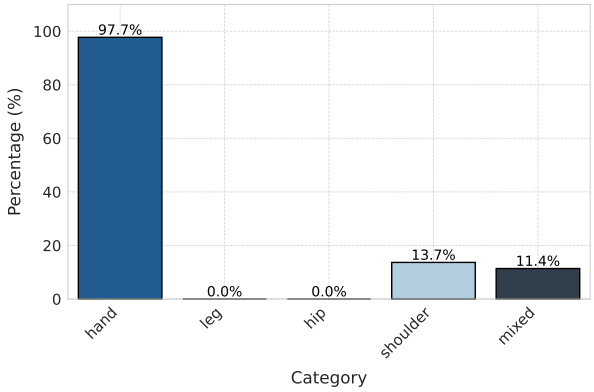

Figure 1. Class distribution in the selected FracAtlas subset.

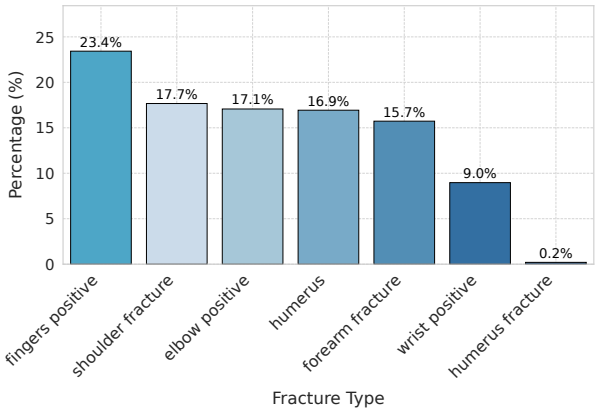

Figure 2. Class distribution in the selected Bone Fracture Detection subset.

The GRAZPEDWRI-DX dataset followed the same pipeline. From $N_{\text{init}} = 3,000$ images, 2,017 valid image–mask pairs were retained (Table II). The resulting subset, with harmonized images, masks, and metadata, is ready for integration with the other datasets, providing a robust training set for fracture detection and segmentation [4][7][18].

Merging these three datasets produced a unified, coherent X-ray collection for upper-limb fracture analysis. Despite differences in acquisition protocols, demographics, and annotation styles, preprocessing, normalization, and validation ensured alignment in image quality, mask formats, and metadata.

After resizing and normalization, summary statistics (Table II) and histograms (Figure 3) highlight variability and

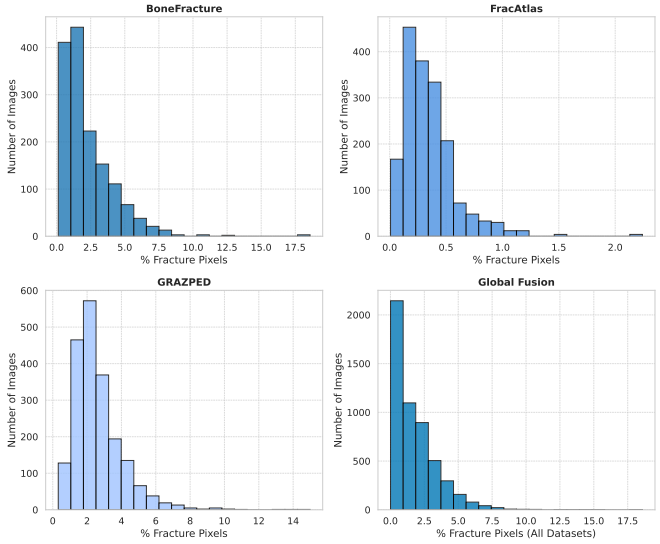

Figure 3. Distribution of fractured pixels across datasets.

potential outliers. Fractures were categorized by size: subtle (0–2% of pixels), medium (2–5%), and large (>5%). Bone fracture detection (Kaggle mainly contains small fractures, FracAtlas mostly subtle ones, and GRAZPED shows a balanced distribution of small to medium fractures, including a few extreme case (Figure 3). Overall, most images feature very small fractures, with larger ones increasingly rare. This distribution suggests a bias toward detecting minor cracks, emphasizing the importance of data augmentation, oversampling, or class-weighted training.

Fracture complexity was quantified using connected components, aspect ratio, and compactness, with the mean complexity across components defining each image's overall complexity. Complexity scores were computed for all images to enable cross-dataset comparisons, and Table II reports mean and maximum values. Notably, small fractures are not always simple; some remain regular while others, though subtle, show intricate shapes, multiple fragments, or irregular boundaries. A full assessment should therefore include fine, medium, and large cases to capture patterns important for robust detection.

The choice of $A = 3$ augmentations per image was subjective, as no data-driven method determined the optimal number. This is a limitation, and future work could test different augmentation levels to evaluate their impact on performance. Remaining limitations also include class imbalance, with subtle fractures being overrepresented, potentially biasing models toward minor fractures. Strategies, such as additional augmentation, synthetic image generation, or class-weighted training are recommended to mitigate this issue.

Moreover, it is important to emphasize that this work is part of an ongoing improvement process. At present, the study relies on publicly available datasets that are annotated for segmentation, contain upper limb X-ray images, and meet our quality standards. The over-representation of certain fracture types, such as wrist fractures, is partly due to their higher prevalence in the population and their greater availability in clinical records, as highlighted by Ye et al. [24]. In the future, if additional datasets meeting these requirements become accessible, they will be incorporated to further increase data diversity, reduce potential bias, and enhance the robustness and generalization capability of the proposed module.

Finally, both the datasets (before and after fusion) and the associated preprocessing, augmentation, and mask generation code are publicly available on the Synapse website, ensuring full reproducibility and enabling task-specific customization. This is made possible through Synapse, a secure platform for sharing, storing, and collaboratively managing biomedical and clinical research data, which allows researchers to access and reproduce datasets reliably [25].

IV. Conclusion and Future Work

This paper introduces a unified and standardized X-ray dataset focused on upper-limb fractures, created by merging FracAtlas, Bone Fracture Detection (Kaggle), and GRAZPEDWRI-DX. The integration pipeline harmonizes image formats, annotations, preprocessing steps, and quality assurance measures, resulting in a consistent and clinically relevant resource suitable for training and evaluating deep learning models. The dataset offers broad anatomical diversity and detailed segmentation masks, supporting research in fracture detection, localization, and segmentation. Remaining challenges include class imbalance and limited representation of complex fractures, which will be addressed through dataset expansion and targeted sampling strategies. Future work will benchmark state-of-the-art architectures and explore clinical deployment potential. Overall, this dataset provides a solid foundation for developing more reliable and generalizable AI-based fracture assessment systems.

Acknowledgment

The authors gratefully acknowledge the financial support provided by Mitacs through the Mitacs Globalink program, which enabled this research internship, as well as by the Fondation Docteur Sadok Besrour.

Data Availability Statement

The unified dataset generated in this study are publicly available in synapse at the following DOI: https://doi.org/10.7303/syn71834100

References

[1] Z. Su, A. Adam, M. F. Nasrudin, M. Ayob, and G. Punganan, "Skeletal fracture detection with deep learning: A comprehensive review," *Diagnostics*, vol. 13, no. 20, p. 3245, 2023. [Online]. Available: https://www.mdpi.com/2075-4418/13/20/3245.

[2] O. Khalilzadeh, C. Canella, and L. M. Fayad, "Wrist and hand," in *Musculoskeletal Diseases 2021–2024: Diagnostic Imaging*, ser. IDKD Springer Series, J. Hodler, R. A. Kubik-Huch, and G. K. von Schulthess, Eds., Springer, Cham, Switzerland, 2021, pp. 41–55. DOI: 10.1007/978-3-030-71281-5_4. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-71281-5_4.

[3] K. Liu et al., "Time to surgical management of distal radius fractures: Effects on health care utilization and functional outcomes," *Canadian Journal of Surgery*, vol. 67, no. 4, E286–E294, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11233172/.

[4] A. Hassan, I. Afzaal, N. Muneeb, A. Batool, and H. Noor, "Ai-based applied innovation for fracture detection in x-rays using custom cnn and transfer learning models," *arXiv e-prints*, arXiv:2509, 2025. [Online]. Available: hhttps://arxiv.org/html/2509.06228v1.

[5] X. Zhai et al., "High-speed x-ray visualization of dynamic crack initiation and propagation in bone," *Acta Biomaterialia*, vol. 90, pp. 278–286, 2019. [Online]. Available: https://doi.org/10.1016/j.actbio.2019.03.045.

[6] J. Ju, Z. Qu, H. Qing, Y. Ding, and L. Peng, "Evaluation of artificial intelligence-based diagnosis for facial fractures, advantages compared with conventional imaging diagnosis: A systematic review and meta-analysis," *BMC Musculoskeletal Disorders*, vol. 26, no. 1, pp. 1–19, 2025. [Online]. Available: https://link.springer.com/article/10.1186/s12891-025-08842-2.

[7] T. Aldhyani et al., "Diagnosis and detection of bone fracture in radiographic images using deep learning approaches," *Frontiers in Medicine*, vol. 11, p. 1 506 686, 2025. [Online]. Available: https://doi.org/10.3389/fmed.2024.1506686.

[8] J. Li et al., "A systematic collection of medical image datasets for deep learning," *ACM Computing Surveys*, vol. 56, no. 5, pp. 1–51, 2023. [Online]. Available: https://dl.acm.org/doi/full/10.1145/3615862.

[9] B. Boecking et al., "Making the most of text semantics to improve biomedical vision–language processing," in *European conference on computer vision*, Springer, 2022, pp. 1–21. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-20059-5_1.

[10] I. Abedeen et al., "Fracatlas: A dataset for fracture classification, localization and segmentation of musculoskeletal radiographs," *Scientific data*, vol. 10, no. 1, p. 521, 2023. [Online]. Available: https://www.nature.com/articles/s41597-023-02432-4.

[16] M. H. Tasin, *Fracatlas original dataset*, [retrieved: January, 2026], 2023. [Online]. Available: https://www.kaggle.com/datasets/mahmudulhasantasin/fracatlas-original-dataset.

[17] P. K. Darabi, *Bone fracture detection: Computer vision project*, [retrieved: January, 2026], 2024. [Online]. Available: https://www.kaggle.com/datasets/pkdarabi/bone-fracture-detection-computer-vision-project.

[18] E. Nagy, M. Janisch, F. Hržić, E. Sorantin, and S. Tschauner, "A pediatric wrist trauma x-ray dataset (grazpedwri-dx) for machine learning," *Scientific data*, vol. 9, no. 1, p. 222, 2022. [Online]. Available: https://www.nature.com/articles/s41597-022-01328-z.

[11] H. M. Lin et al., "The rsna cervical spine fracture ct dataset," *Radiology: Artificial Intelligence*, vol. 5, no. 5, e230034, 2023.

[12] D. D. Ruikar, K. Santosh, R. S. Hegadi, L. Rupnar, and V. A. Choudhary, "5k+ ct images on fractured limbs: A dataset for medical imaging research," *Journal of Medical Systems*, vol. 45, no. 4, p. 51, 2021.

[0] G. Duan, S. Zhang, Y. Shang, and W. Kong, "Research on x-ray diagnosis model of musculoskeletal diseases based on deep learning," *Applied Sciences*, vol. 14, no. 8, p. 3451, 2024.

[13] P. Rajpurkar et al., "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1711.05225.

[14] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017. [Online]. Available: https://pubs.rsna.org/doi/abs/10.1148/radiol.2017162326.

[15] I. M. Wani and S. Arora, "Computer-aided diagnosis systems for osteoporosis detection: A comprehensive survey," *Medical & biological engineering & computing*, vol. 58, no. 9, pp. 1873–1917, 2020. [Online]. Available: https://link.springer.com/article/10.1007/s11517-020-02171-3.

[19] A. A. Roque and H. Sebastian, "Chest x-ray pneumothorax segmentation using efficientnet-b4 transfer learning in a u-net architecture," *arXiv preprint arXiv:2509.03950*, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.2509.03950.

[20] E. Goceri, "Medical image data augmentation: Techniques, comparisons and interpretations," *Artificial intelligence review*, vol. 56, no. 11, pp. 12 561–12 605, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s10462-023-10453-z.

[21] M. Kutbi, "Artificial intelligence-based applications for bone fracture detection using medical images: A systematic review," *Diagnostics*, vol. 14, no. 17, p. 1879, 2024. [Online]. Available: https://www.mdpi.com/2075-4418/14/17/1879.

[22] J. A. Parker, R. V. Kenyon, and D. E. Troxel, "Comparison of interpolating methods for image resampling," *IEEE Transactions on medical imaging*, vol. 2, no. 1, pp. 31–39, 2007. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4307610.

[23] R. Harini and C. Chandrasekar, "Image segmentation using nearest neighbor classifiers based on kernel formation for medical images," in *International conference on pattern recognition, informatics and medical engineering (prime-2012)*, IEEE, 2012, pp. 261–265. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6208355.

[24] J. Ye, Q. Li, and J. Nie, "Prevalence, characteristics, and associated risk factors of wrist fractures in americans above 50: The cross-sectional nhanes study," *Frontiers in endocrinology*, vol. 13, p. 800 129, 2022.

[25] Sage Bionetworks, *Synapse: A collaborative data sharing and research platform*, https://sagebionetworks.org/platform/synapse, [retrieved: January, 2026], 2026.