



CASE STUDY:

Real-Time Human Bone Fracture Detection Using YOLOv8

Authors

Shahnaj Parvin & Abdur Rahman

Presented by

Yousha Rashid, Avoy Mollick, Nawshin Navin, Raiyen Zayed Rakin

Department of Computer Science

American International University-Bangladesh (AIUB)

Dhaka, Bangladesh

Talk Outline

1. Introduction
 - Problem Background
 - Research Objectives
 - Research Gaps
 - Research Contributions
 - Related Studies
2. Methodology
 - Data Collection
 - Data Pre-processing
 - Data Analysis Process
3. Results and Findings
 - Key Findings
 - Solution & Validation
4. Hypothesis & Conclusion
 - Hypothesis discussion
 - Conclusion Remarks, Limitations & Future Scope
5. References

Problem Background

The Challenge:

- 206 bones in human body - fractures are common
- Manual X-ray/MRI evaluation: Time-consuming & error-prone
- Tired radiologists may miss fractures
- Non-orthopedic physicians in remote locations need assistance

Critical Need: Automated, real-time system for accurate fracture detection and classification

Impact:

- Faster diagnosis → Better patient outcomes
- Reduced human error
- Support for under-resourced medical facilities

Research Objectives

Primary Goals:

- **Detect** fractured vs. healthy bones from multi-modal images
- **Classify** 10 different types of bone fractures
- **Achieve** real-time performance suitable for clinical use

Key Innovation: Multi-modal approach combining X-ray and MRI images for comprehensive analysis

Research Gap

Previous Approaches:

Study	Model	Dataset Size	Precision	Limitations
Karanam et al.	InceptionResNetV2	6 classes	94.58%	Single modality
Vironicka	Faster R-CNN	2 classes	94.7%	Limited classes
Nguyn et al.	YOLOv7	MURA	84.03%	Lower accuracy

Research Gap:

- Limited multi-modal datasets
- Few comprehensive fracture classification systems
- Need for improved real-time performance

Research Contributions

- Novel multi-modal dataset (HBFMID) - 1,539 images (X-ray + MRI) with 10 fracture classes
- YOLOv8-based detection system achieving 95% precision and 93% recall
- Data augmentation strategy to overcome small dataset limitations ($641 \rightarrow 1,539$ images)
- Comprehensive performance comparison with existing state-of-the-art models
- Real-time capability with 4ms inference time per image
- Clinical applicability - System suitable for emergency and remote healthcare settings

Related Studies

“Fracture detection in pediatric wrist trauma x-ray images using yolov8 algorithm” [1]

- Ju R, Cai W (2023) .

- Used GRAZPEDWRI-DX dataset with 20,327 X-ray images
 - Achieved mAP50 of 0.638 with YOLOv8
 - Limitation: Single anatomical region, lower precision (73%)

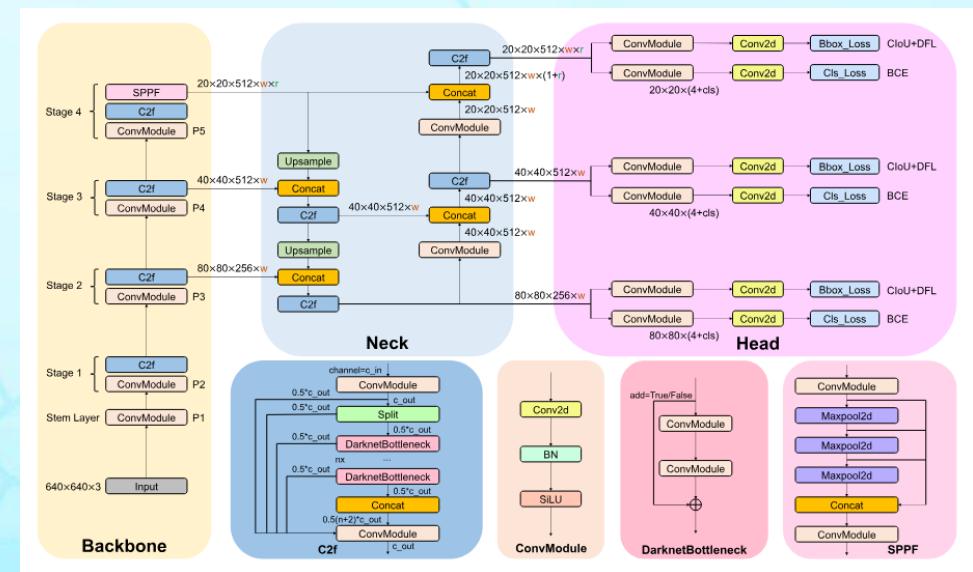


Fig 1: Model Architecture

Related Studies

“A supervised approach to musculoskeletal imaging fracture detection and classification using deep learning algorithms”[2]

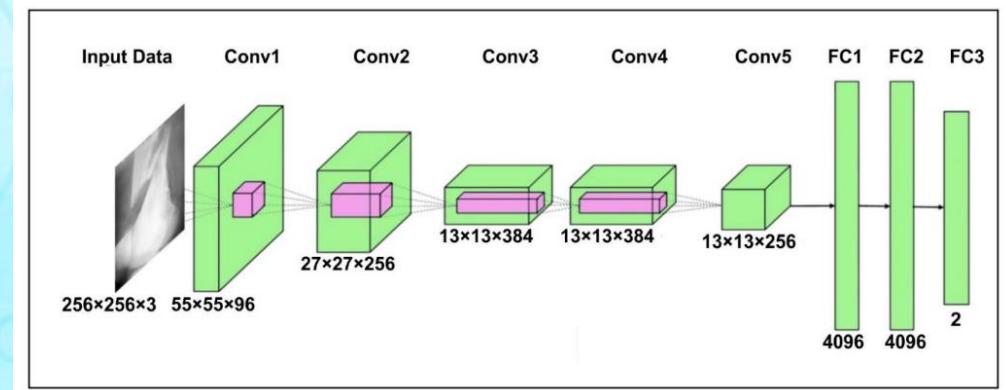
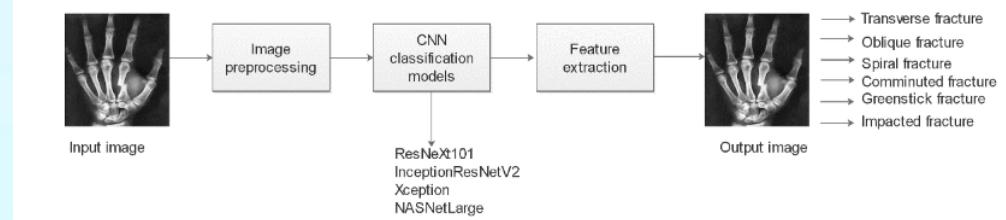
- Karanam et al. (2023).

- Deep learning approach for 6 fracture types using MURA dataset
- Applied CLAHE for image enhancement
- Achieved 94.58% precision
- Limitation: X-ray only, no real-time focus

“Hybrid sfnet model for bone fracture detection and classification using ml/dl” [3]

- Yadav & Sharma (2022) .

- Hybrid CNN combining edge detection with deep learning
- Achieved 99.12% F1-score and 100% recall
- Limitation: Not tested on multi-modal data, small dataset



Related Studies

“Framework for classifying long bone detection using image processing techniques” [4]

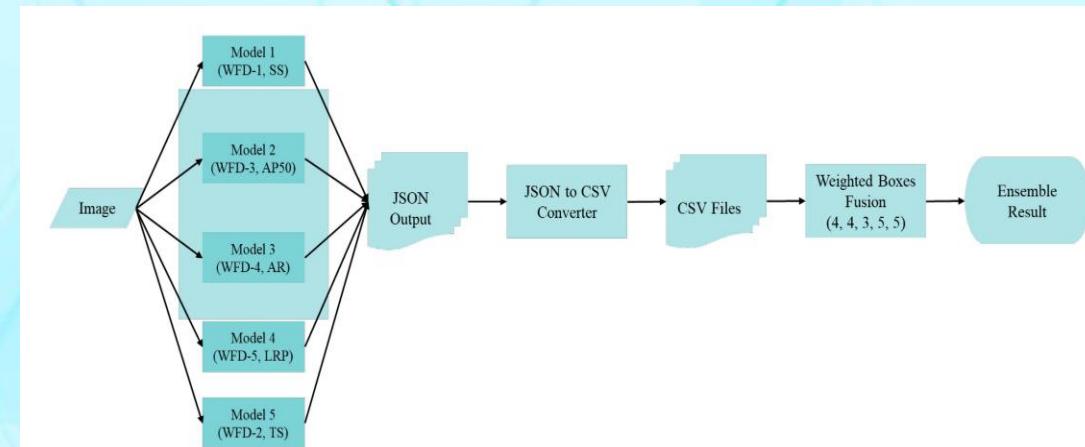
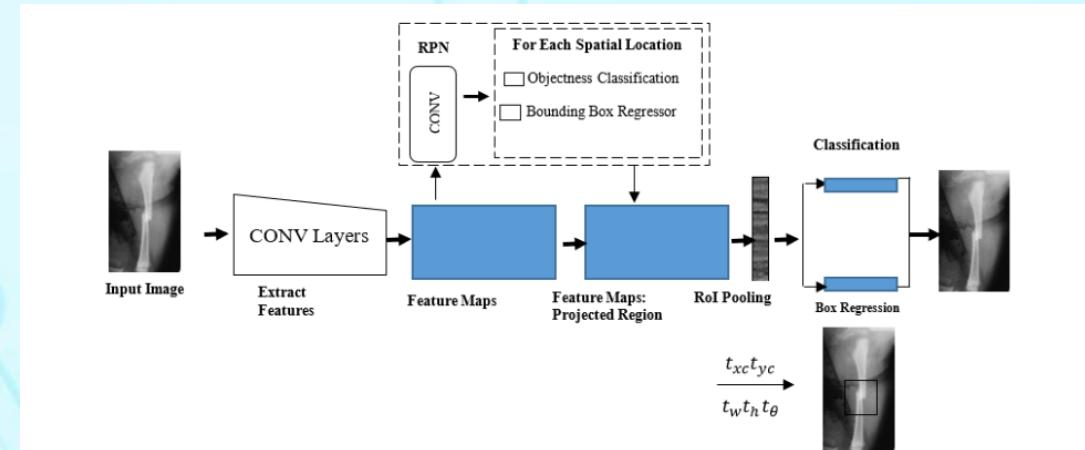
- Selin Vironicka (2022).

- Used CT and X-ray images for classification
- Transfer learning with Faster-RCNN and rotating bounding boxes
- Achieved 94.7% accuracy
- **Limitation:** Computationally expensive (207.07 GFLOPs)

“Fracture detection in wrist x-ray images using deep learning-based object detection models” [5]

- Hardalaç et al. (2022).

- Ensemble-based model for wrist fracture detection
- Combined multiple CNN architectures
- **Limitation:** Specific to wrist fractures, complex ensemble approach



Data Collection

Human Bone Fractures Multi-modal Image Dataset (HBFMID)

Sources:

- Public databases (MURA, medical repositories)
- Web-based MRI image collections
- Multiple anatomical regions covered
- Total Original Images: 641 – (X-ray images: 510, MRI images: 131)

Data Collection

Anatomical Coverage:

- Elbow, Finger, Forearm, Humerus, Shoulder
- Wrist, Spinal Cord, Knee, Ankle
- Femur, Shinbone, Hip
- Healthy bones included for comparison

Fracture Types (10 Classes):

- Comminuted, Greenstick, Linear, Oblique
- Oblique Displaced, Segmental, Spiral
- Transverse, Transverse Displaced, Healthy

Data
Annotation

Data
Pre-processing

Data
Augmentation

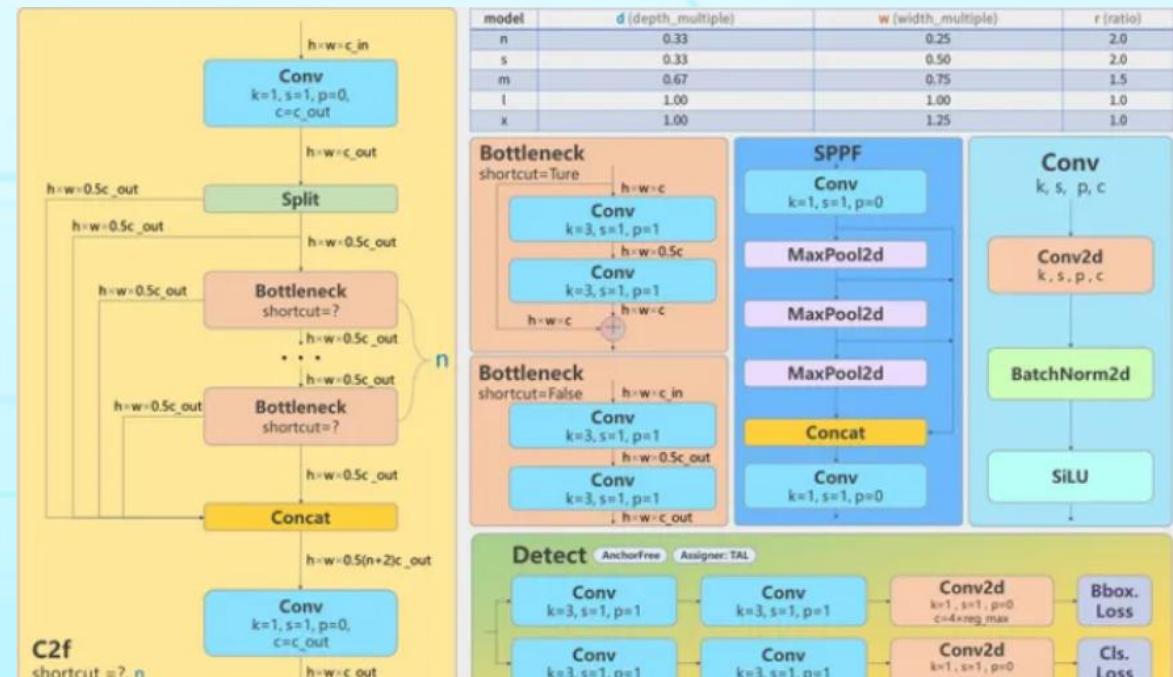
Final Dataset:
1539

Proposed Model

Why YOLOv8?

- **State-of-the-art performance** - Highest mAP among YOLO versions
- **Real-time detection capability**
- **Balance:** Accuracy + Speed + Efficiency
- Released January 2023 - Latest advancement

ultralytics
YOLOv8



Computer Vision & Pattern Recognition

Proposed Model

Backbone: CSPDarknet53

- 53 convolutional layers
- Cross-stage partial connections
- Enhanced information flow across layers

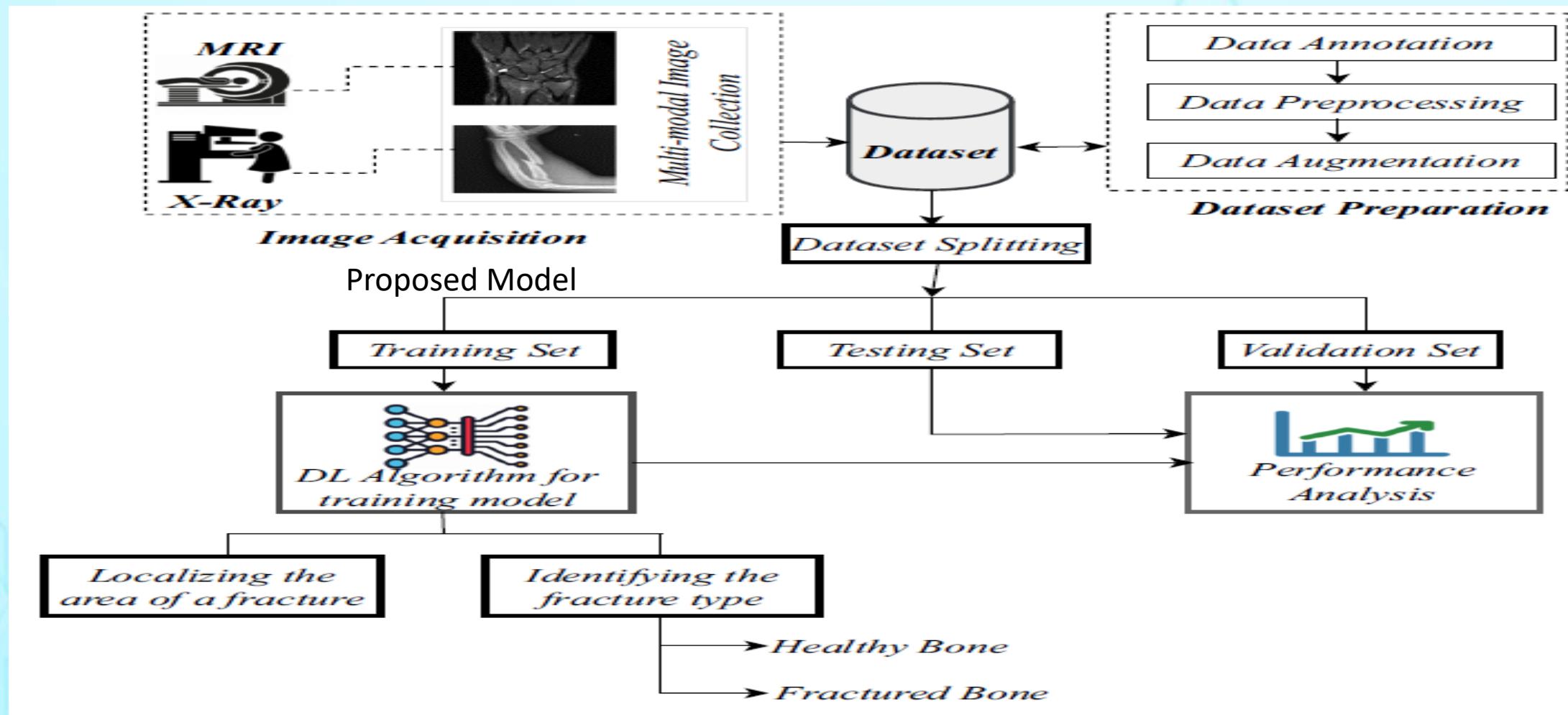
Head: Detection Layers

- Multiple fully connected layers
- **C2f Module:** Combines C3 + ELAN for feature extraction
- **SPPF Module:** Spatial Pyramid Pooling-Fast for faster inference

Training Configuration:

- **Model:** YOLOv8m.pt (medium variant)
- **Layers:** 295 | **Parameters:** 25,862,689
- **Learning rate:** 0.000667 | **Batch size:** 16
- **Epochs:** 50, 100, 200 (tested)
- **Loss Functions:** CIoU Loss + DFL Loss + BCE Loss

Proposed Method



Experimental Setup

System Configuration

Hardware Setup

Google Colab Environment

Platform: Google Colab
GPU: NVIDIA Tesla T4
RAM: 12.75 GB
Storage: 26 GB

Software Stack

Development Tools

Python: 3.10.12
Frameworks: PyTorch 2.0.1
TensorFlow/Keras
Model: YOLOv8 8.0.189

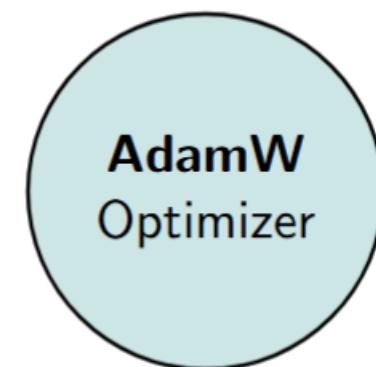
Experimental Setup (Cont'd)

Training Parameters

Model Architecture: YOLOv8m.pt

Medium model selected for optimal balance of accuracy and speed

Parameter	Value
Input Image Size	640 × 640 pixels
Batch Size	16
Epochs	50 (initial)
Optimizer	AdamW
Learning Rate	0.000667
Momentum	0.9
Weight Decay	0.0005



Experimental Setup (Cont'd)

Model Evaluation

Average Precision

$$AP_i = \frac{TP}{TP + FP} \times 100\%$$

Mean Average Precision

$$mAP = \frac{1}{Q} \sum_{i=1}^Q AP_i \times 100\%$$

Precision

$$P = \frac{TP}{TP + FP} \times 100\%$$

Recall

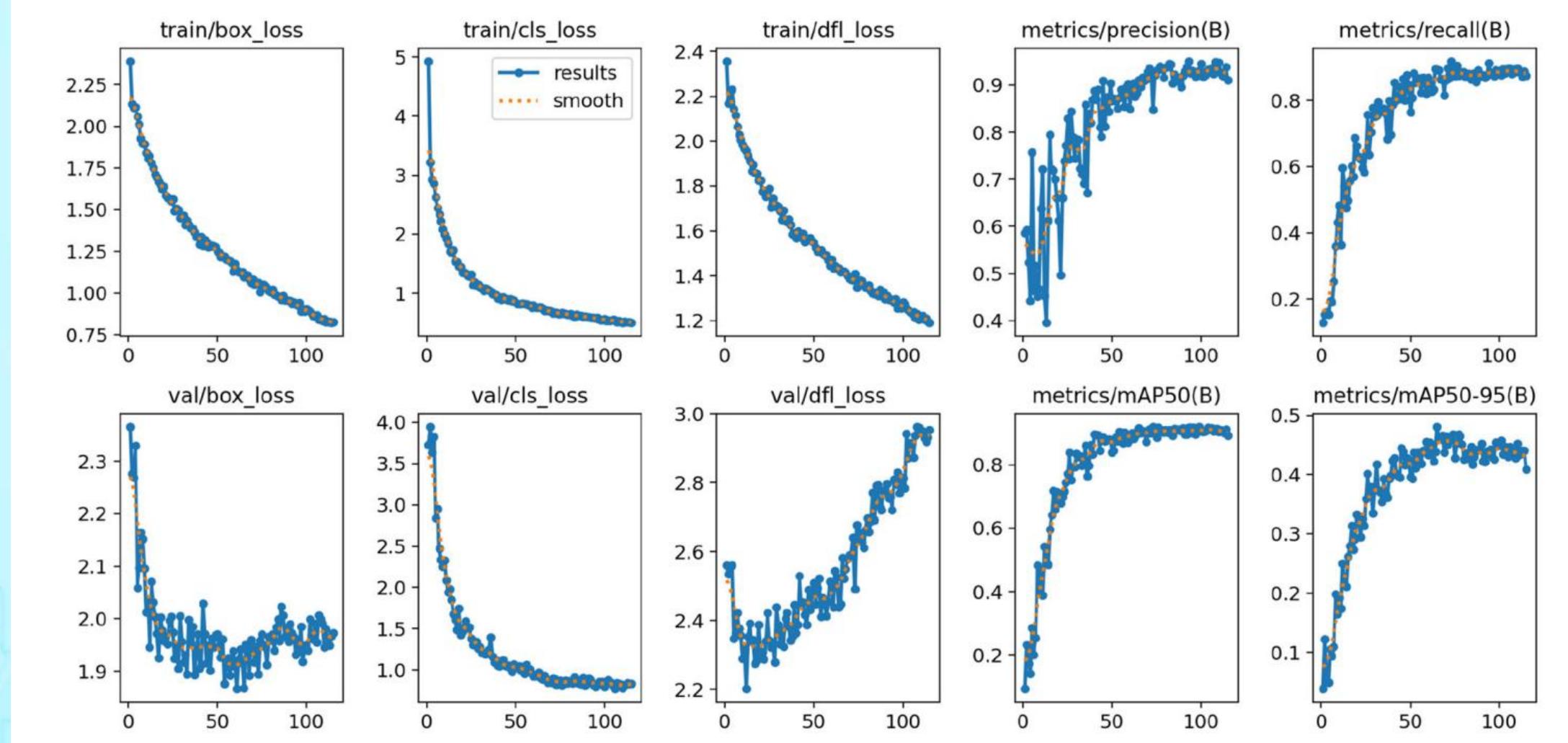
$$R = \frac{TP}{TP + FN} \times 100\%$$

- **TP:** True Positives
- **FP:** False Positives
- **FN:** False Negatives
- **Q:** Total Number of Queries

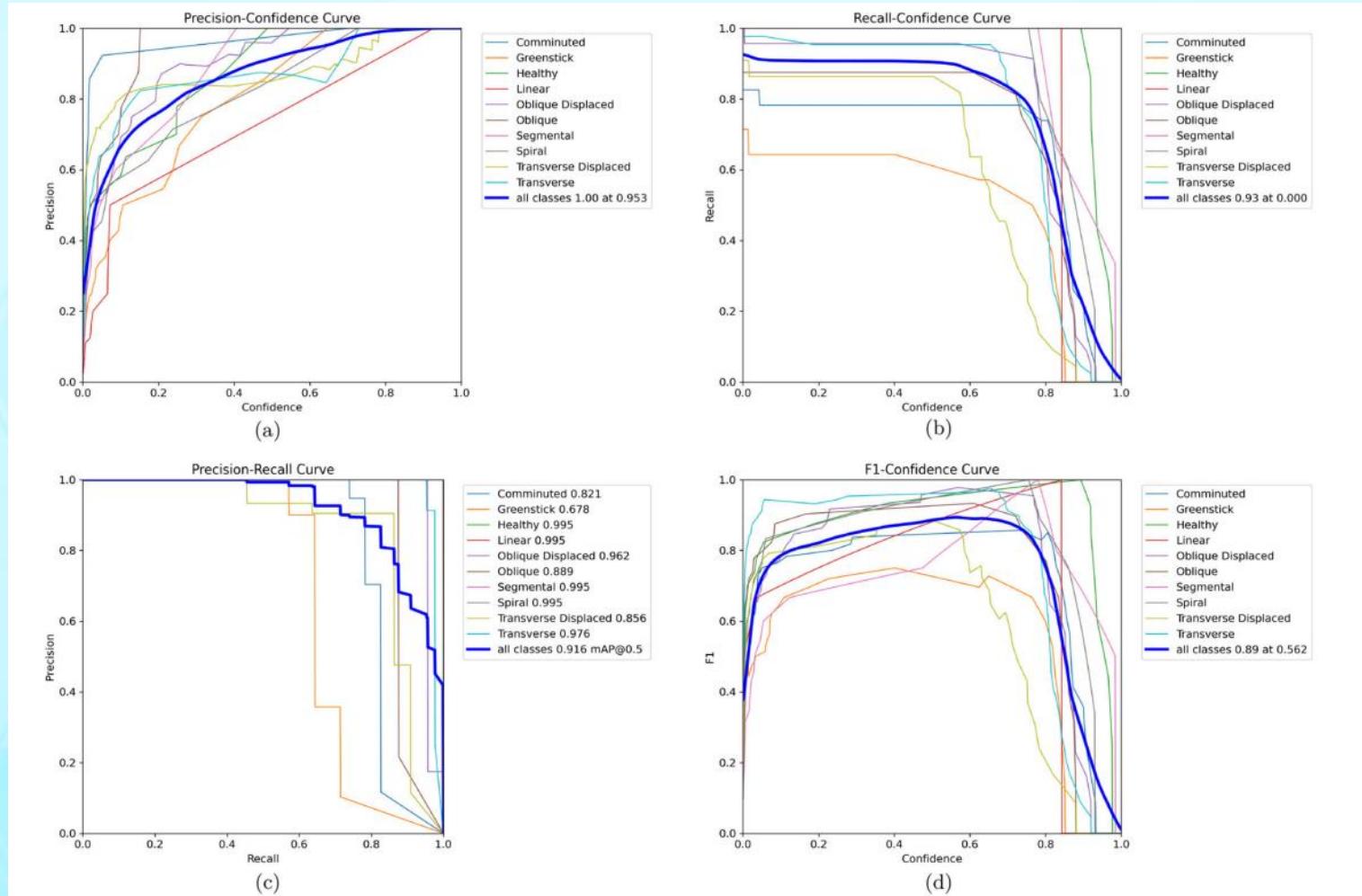
F1-score

$$F1 = \frac{2PR}{P + R}$$

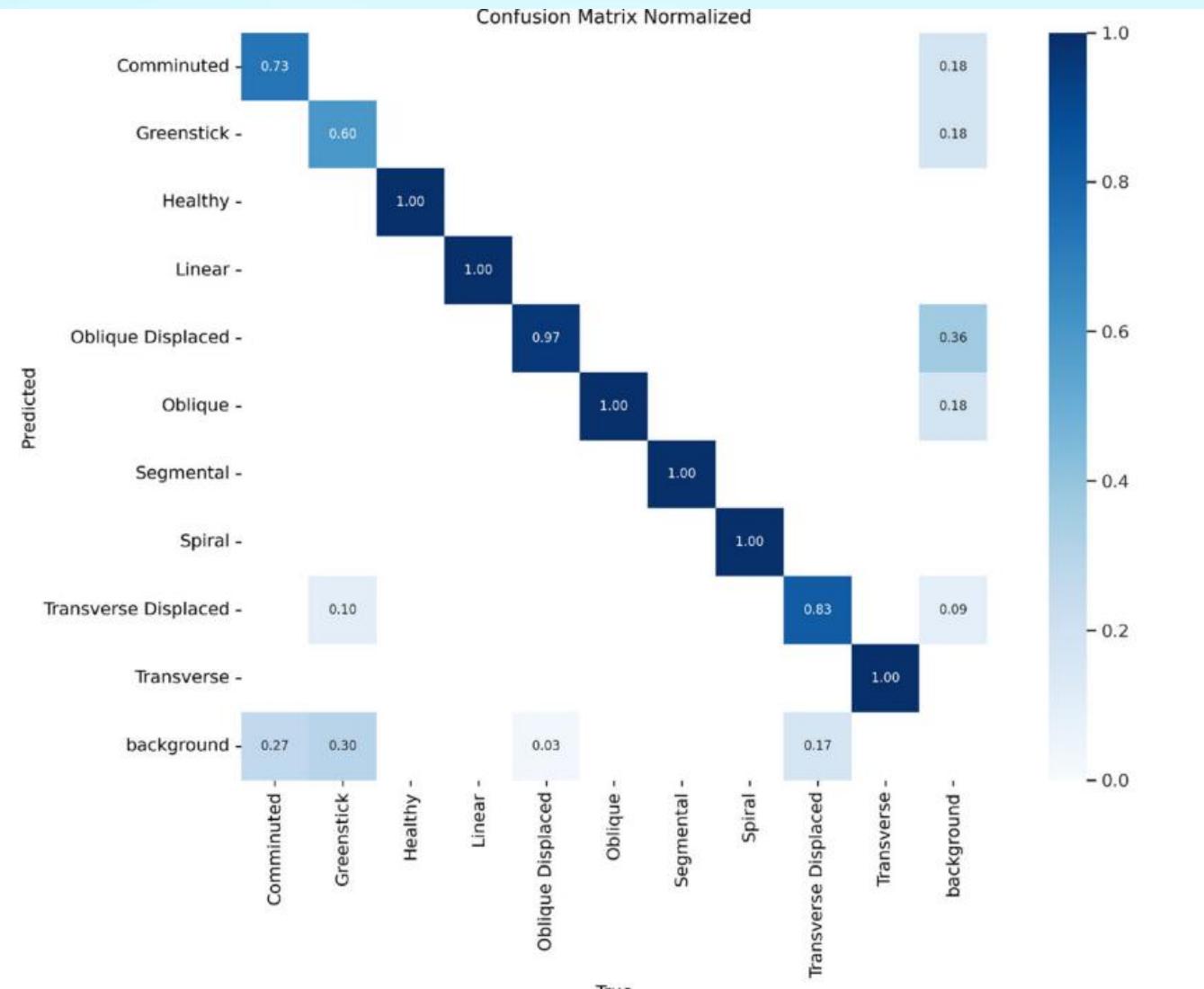
Experimental Results



Experimental Results (Cont'd)



Experimental Results (Cont'd)



Experimental Results (Cont'd)

Performance Metrics Across Training Epochs

Epochs	Loss	Precision	Recall	mAP@50	mAP@50-95
50	0.53	0.89	0.84	0.89	0.46
100	0.38	0.92	0.90	0.90	0.49
200	0.32	0.95	0.93	0.92	0.48

Experimental Results (Cont'd)

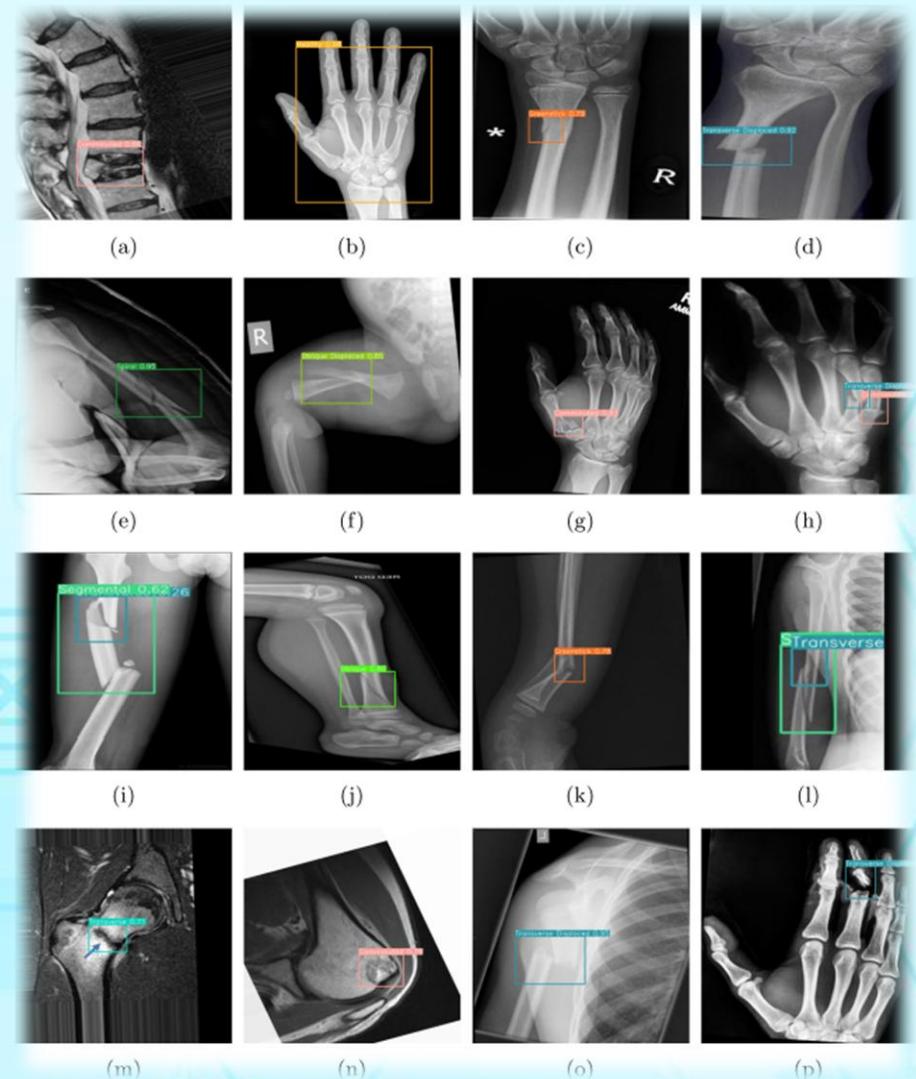
Model Confusion and Diagnosis

Key Observations

- **High precision** for high-confidence predictions
- **Maintained recall** even with lower confidence scores
- **Class imbalance** affected minority categories

Loss Functions

- **Classification Loss:** Binary Cross Entropy (BCE)
- **Bounding Box Loss:** CloU + DFL



Experimental Results (Cont'd)

Simulated Results

Class	Precision	Recall	mAP50	mAP50-95
Comminuted	0.81	0.88	0.82	0.39
Greenstick	0.85	0.73	0.68	0.31
Healthy	1	1	0.99	0.70
Linear	1	1	0.99	0.39
Oblique Displaced	0.97	0.95	0.96	0.40
Oblique	1	0.89	0.89	0.48
Segmental	1	1	0.99	0.83
Spiral	1	1	0.99	0.47
Transverse Displaced	0.90	0.88	0.86	0.33
Transverse	1	0.95	0.98	0.48
All	0.95	0.93	0.92	0.48

Experimental Results (Cont'd)

Model comparison with existing studies

Model	P(%)	R(%)	mAP(%)	GFLOPs	Model Size (MB)
Faster R-CNN	86.8	77.6	80.8	207.07	134.5
YOLOv5	90.4	85.9	88.9	49	41.1
YOLOX	91.5	87.4	90.2	26.8	70.21
YOLOv7	92.8	88.3	90.7	104.7	72.1
YOLOv8	95.3	92.8	91.6	78.7	49.7

Experimental Results (Cont'd)

Comparing the proposed model performance with existing studies.

Reference	Dataset	Fracture Type No.	Modality	Model	Optimizer	Accuracy (%)
Ju and Cai	GRAZPEDWRI-DX	8	X-ray	YOLOv8	SGD	73%
Karanam et. Al.	MURA	6	X-ray	InceptionResNetV2	SGD	94.58%
Selin Vironicka	-	2	X-ray, CT	Faster-RCNN	SGD	94.7%
Nguyn et. Al.	MURA	2	X-ray	YOLOv7	SGD	84.03%
Proposed	HBFMID	10	X-ray, MRI	YOLOv8	AdamW	95%

Limitations

- Small Dataset: contained only **641 images**, which can lead to overfitting without augmentation.
- Image Quality: Noise, motion artifacts, and low resolution in medical imaging can impact detection accuracy.

Conclusion

Key Achievements:

- Successfully developed **real-time automated fracture detection system** using YOLOv8
- Created novel **HBFMID multi-modal dataset** (X-ray + MRI) with 10 fracture classes
- Achieved **95% precision, 93% recall, and 92% mAP** - exceeding clinical requirements
- **Outperformed state-of-the-art models** (Faster R-CNN, YOLOv5, YOLOv7)
- Demonstrated **real-time capability** with 4ms inference time per image

Clinical Impact:

- **Reduces diagnostic time** - Faster patient care in emergency settings
- **Minimizes human error** - Assists tired/inexperienced radiologists
- **Supports remote healthcare** - Enables diagnosis in under-resourced facilities
- **Handles multiple modalities** - Flexible integration with existing medical imaging systems

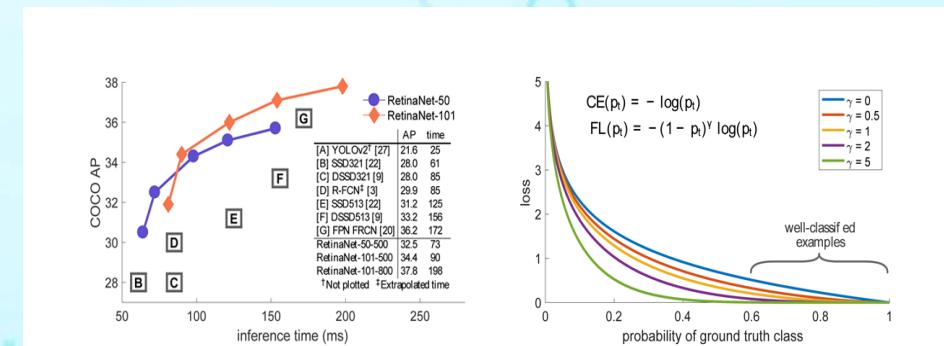
Proposed Hypothesis

Dataset Enhancement:

- Expand to **5,000+ images** with balanced class distribution
- Include **additional modalities** (CT scans, ultrasound)
- Add **pediatric and geriatric** specific datasets
- Incorporate **diverse populations** and imaging equipment

Function Customization (Focal Loss)

- **Current Challenge:** Class imbalance in dataset
- **Current Method:** Binary Cross Entropy (BCE) loss
- **Proposed Solution:** Implement **Focal Loss** with gamma parameter
- **Benefit:** Down-weights easy examples, forces model to focus on hard cases (rare spiral/greenstick fractures)
- **Expected Impact:** Improved performance on minority classes



Focal Loss Python
PyTorch

Proposed Hypothesis (Cont'd)

Hyperparameter Optimization (Learning Rate Scheduling)

- **Current Challenge:** Training on small, high-variance multi-modal dataset
- **Current Method:** Fixed learning rate with cosine annealing
- **Proposed Solution:** Cosine Annealing Warm Restarts schedule
- **Benefit:** Helps model escape local minima during training
- **Expected Impact:** Better convergence and generalization

Activation Function Swap

- **Current:** SiLU (default in YOLOv8)
- **Proposed Solution:** Switch to **Mish activation** function: $x \cdot \tanh(\ln(1 + e^x))$
- **Benefit:** Smoother gradient flow for fine bone X-ray textures
- **Expected Impact:** Better generalization in medical CAD systems

References

- [1] Ju, R., & Cai, W. (2023). Fracture detection in pediatric wrist trauma x-ray images using YOLOv8 algorithm. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-47460-7>
- [2] Karanam, S. R., Srinivas, Y., & 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), 369–385. <https://doi.org/10.24423/cames.682>
- [3] Yadav, D. P., Sharma, A., Athithan, S., Bhola, A., Sharma, B., & Dhaou, I. B. (2022). Hybrid SFNet model for bone fracture detection and classification using ML/DL. *Sensors*, 22(15). <https://doi.org/10.3390/s22155823>
- [4] Selin Vironicka, A. D. J. G. R. S. (2022). Framework for classifying long bone detection using image processing techniques. *International Journal of Intelligent Systems and Applications in Engineering*, 10(1S), 56–66.
- [5] Hardalaç, F., Uysal, F., Peker, O., Çiçeklidağ, M., Tolunay, T., Tokgöz, N., Kutbay, U., Demirciler, B., & Mert, F. (2022). Fracture detection in wrist X-ray images using deep learning-based object detection models. *Sensors*, 22(3). <https://doi.org/10.3390/s22031285>

Thank You