Severity-Aware Multimodal Network for Chest X-Ray Triage and Disease Captioning

Zinah Ghulam[1], Richa Mittal[2], Eranga Ukwatta[1]

[1]College of Engineering, University of Guelph, 50 Stone Rd E, Guelph, ON, Canada
[2]Radiology Department, Guelph General Hospital, Guelph, ON, Canada

INTRODUCTION

The Problem: Chest X-ray (CXR) interpretation requires specialized expertise and is time-consuming, causing delays that impact patient care [1].

Prior Work: Deep learning models excel at classification [1, 2], but multimodal approaches better integrate images with textual clinical context [3].

MOTIVATION

The Gap: Automated triage based on disease urgency remains underexplored [4]. Existing systems rarely combine interpretable captioning with continuous severity scoring [5, 6].

OBJECTIVE

To develop a **Multimodal Triage Network (MTN)** that jointly performs:

1. Interpretable disease captioning.
2. Continuous severity scoring (0.0-3.0 scale).
3. Efficient patient prioritization and real-time triage.

CLINICAL GROUND TRUTH DEVELOPMENT

We fused categorical labels with linguistic modifiers (Fig. 1), to create a 0.0-3.0 scale. Using BioBERT and the rules in Table 1, the model identifies "acute" versus "stable" findings to prioritize worsening conditions.
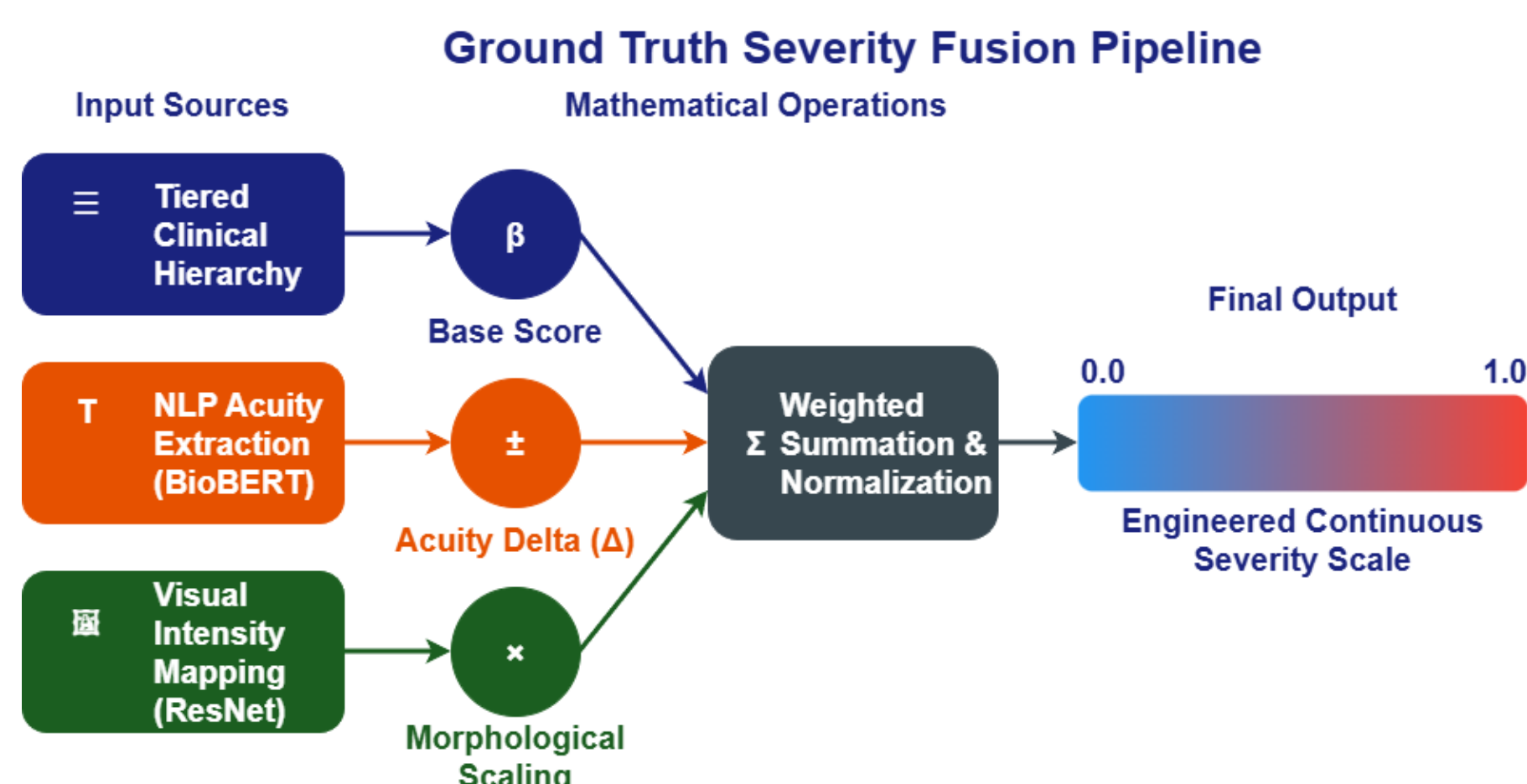

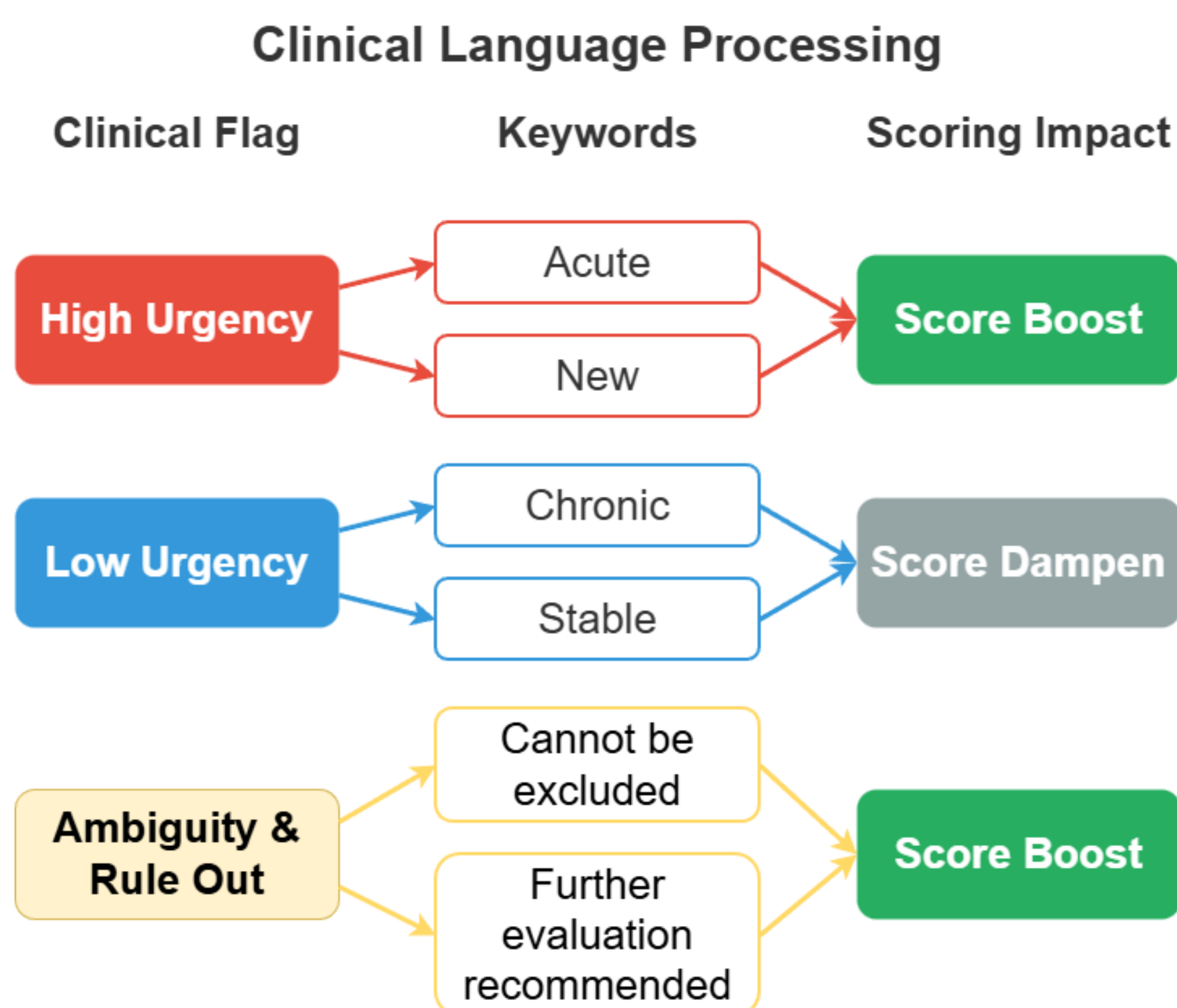**Fig. 1.** Ground Truth Design (Clinical Influence Factors)

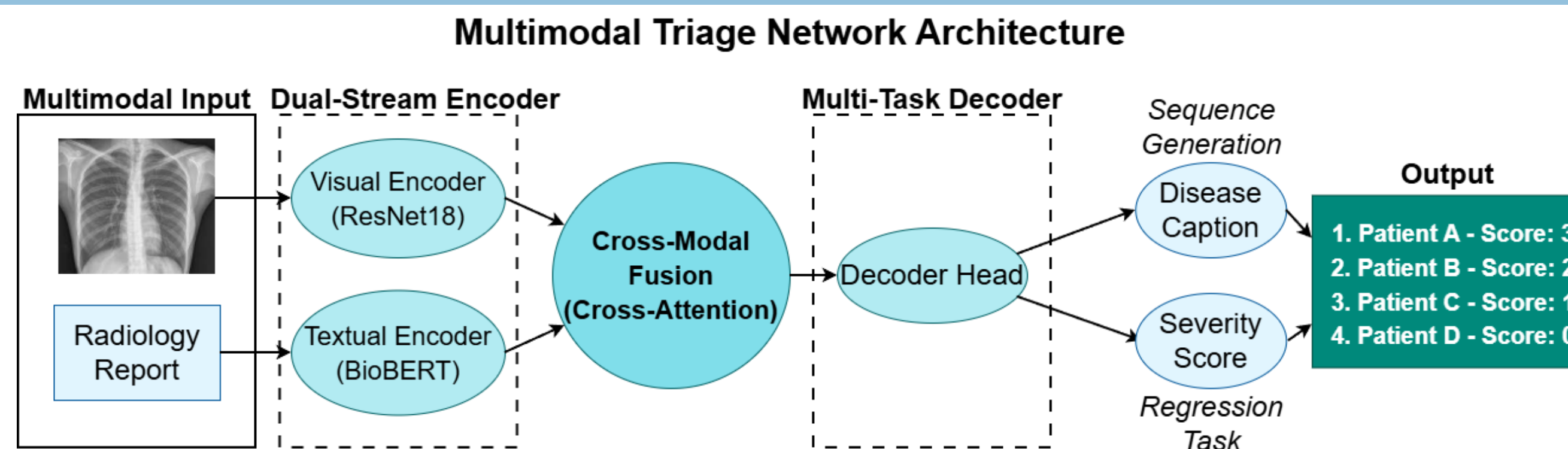**Fig. 2:** Clinical Language Processing Rules

METHODS

Multimodal Triage Network Architecture

Fig. 3: Overall architecture, consisting of a dual-stream encoder, cross-modal fusion, and multi-task decoder.

Dual-Stream Encoder Architecture

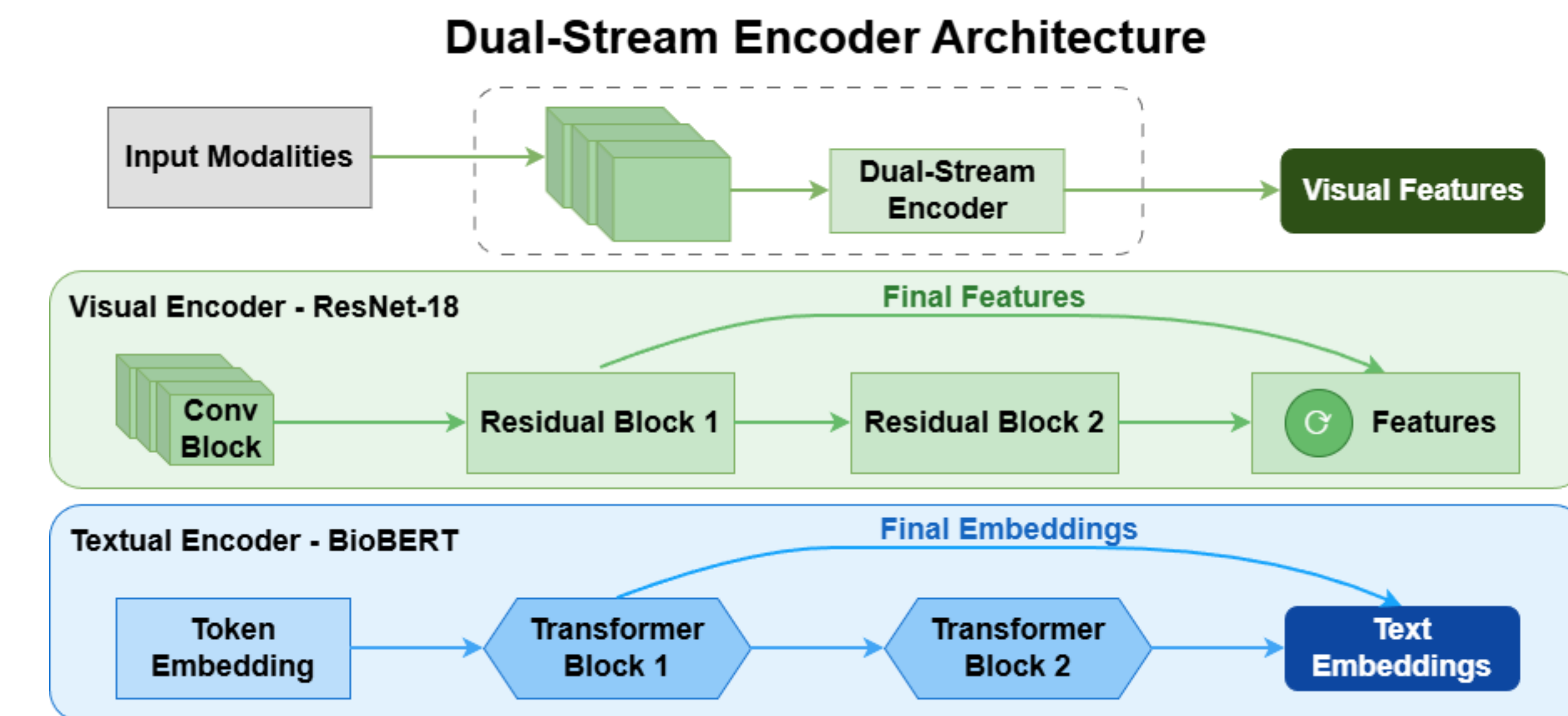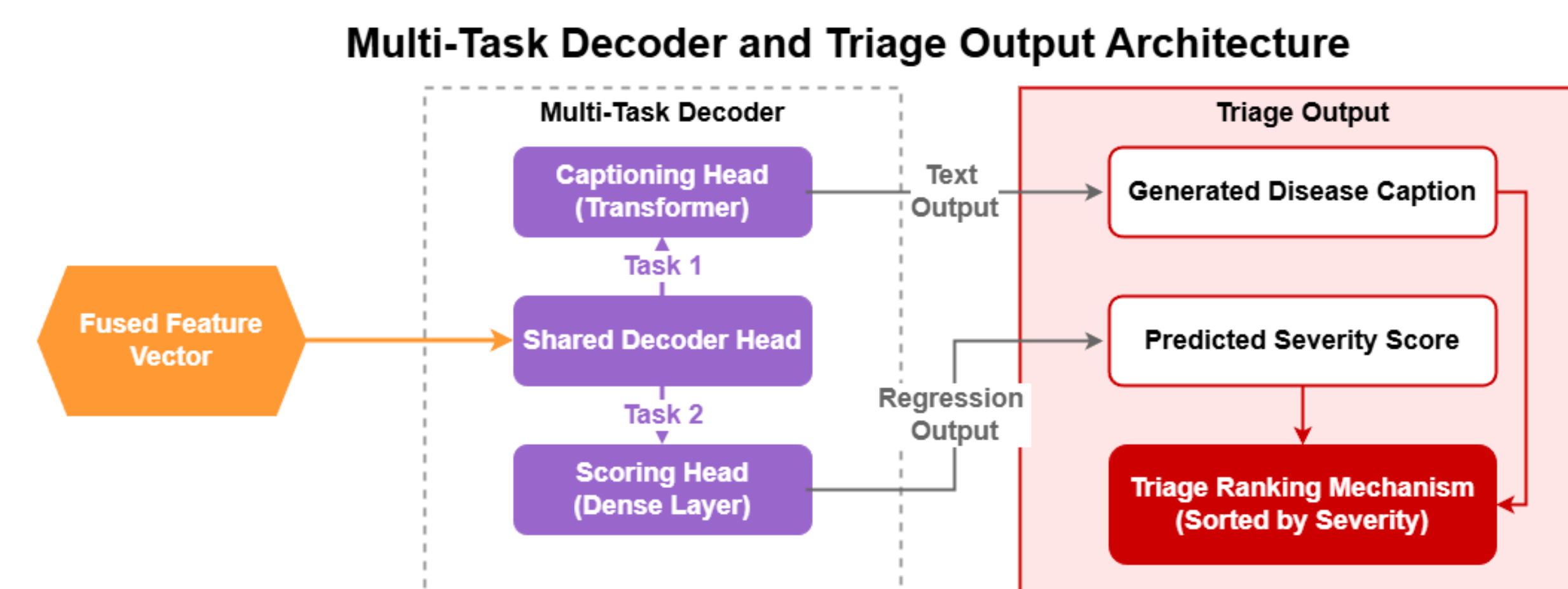**Fig. 4:** Multi-task decoder architecture.

Multi-Task Decoder and Triage Output Architecture

Fig. 5: Multi-task decoder architecture.

Ground Truth Design:
Table 1: Severity Tiers and Clinical Mapping

Priority Tier	Descriptive Label	Conditions
Tier 3	Critical / High Urgency	Pneumothorax, Edema, Pneumonia, Consolidation, Lung Opacity, Atelectasis, Enlarged Cardiomediastinum (in trauma), Misplaced Support Devices.
Tier 2	Important / Moderate	Fracture, Lung Lesion (Size-dependent), Pleural Other.
Tier 1	Low Urgency	Pleural Effusion
Tier 0	Not Urgent	Cardiomegaly, Enlarged Cardiomediastinum (non-trauma), Properly placed devices, No Finding.

RESULTS

Ablation Study:

Table 2. Performance comparison of MTN and unimodal baselines.

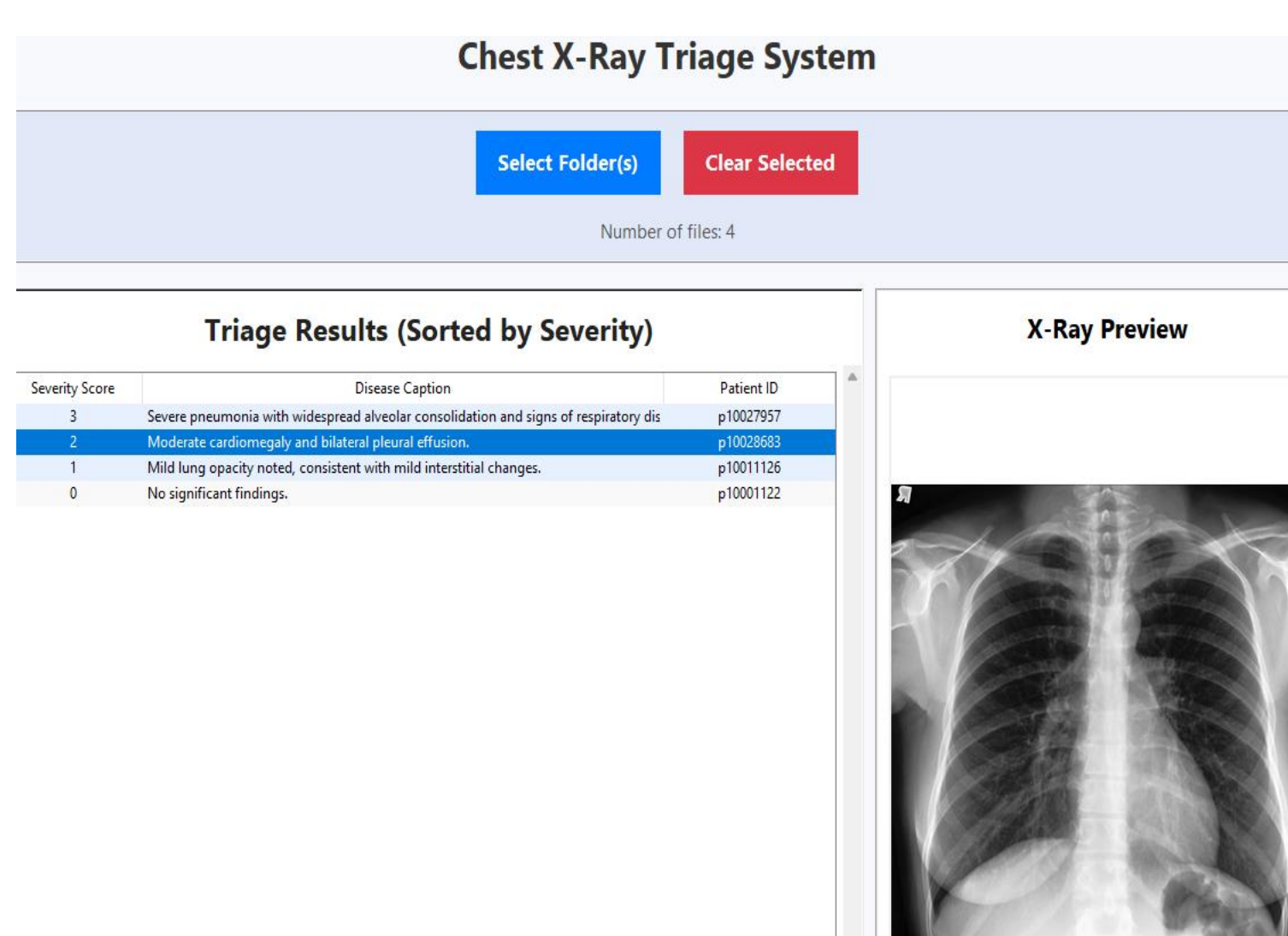| Model	F1-score	MSE
Visual-only (ResNet18)	0.8500	0.1500
Text-only (BioBERT)	0.5100	0.8600
MTN (Ours)	**0.9746**	**0.0448**

Fig. 6: MTN interface displaying ranked patient cases with captions and severity scores; critical cases are prioritized for rapid triage.
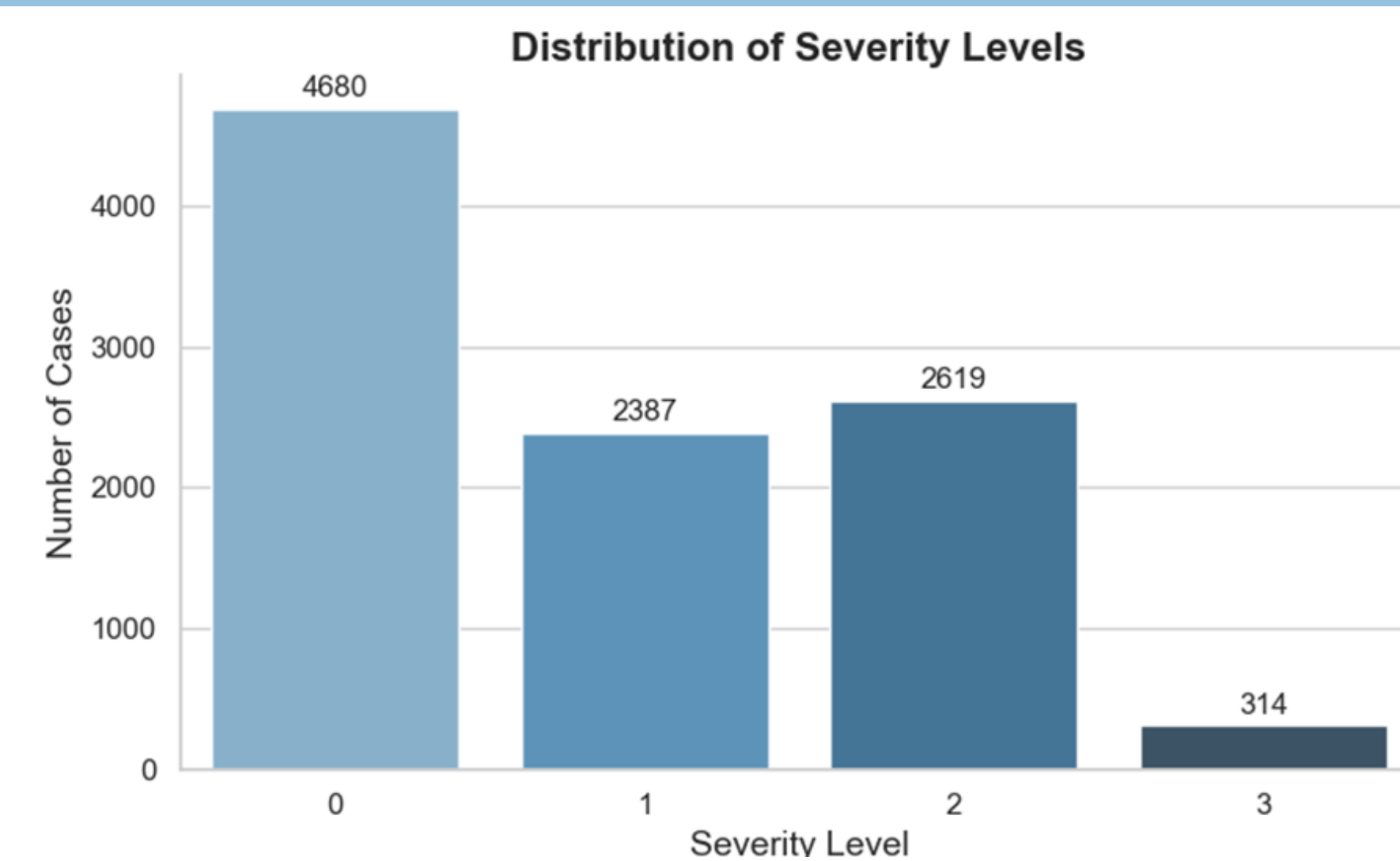
EVALUATION

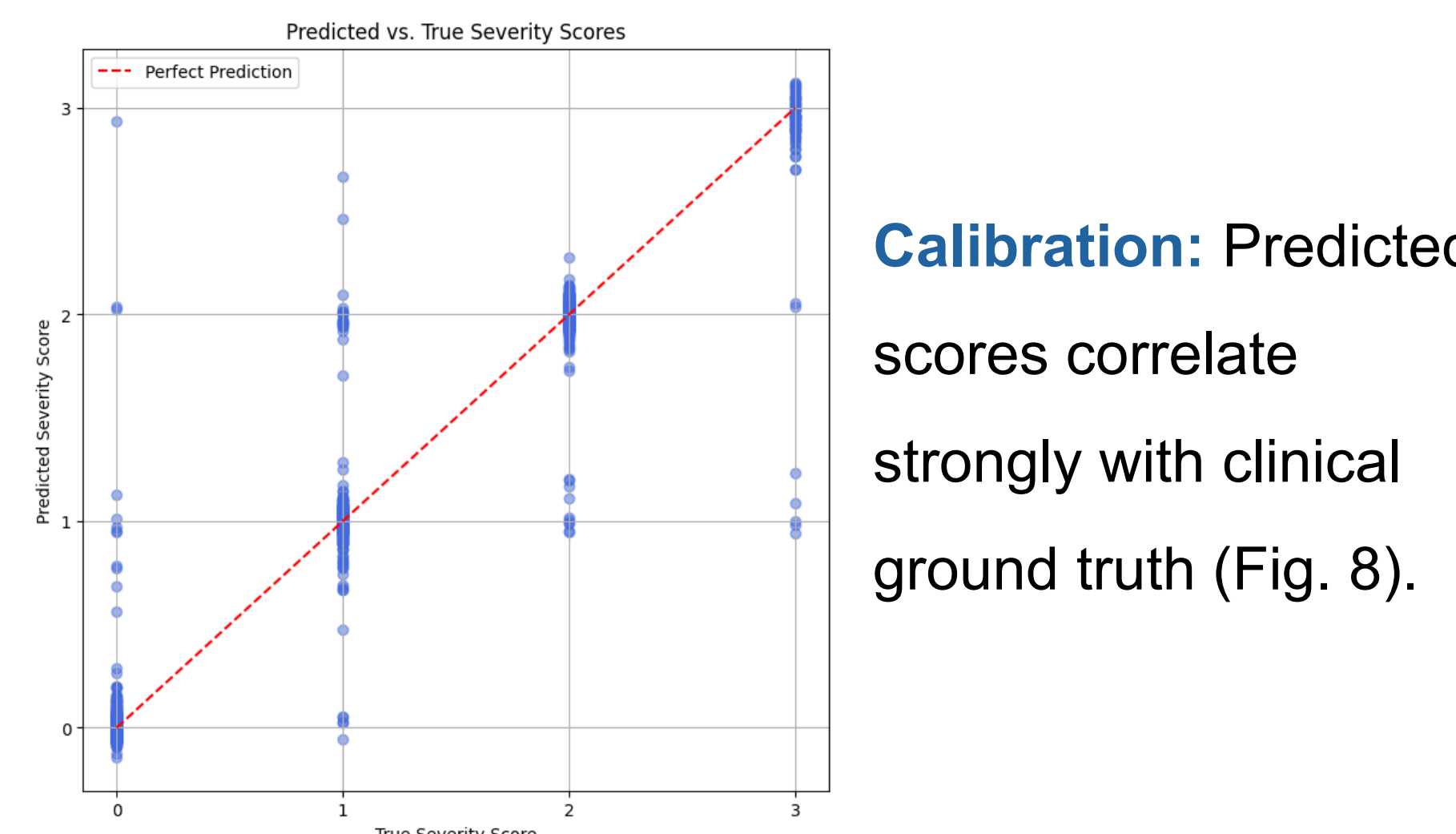**Fig. 7:** Distribution of disease severity levels (0-3) in the curated dataset.

Calibration: Predicted scores correlate strongly with clinical ground truth (Fig. 8).

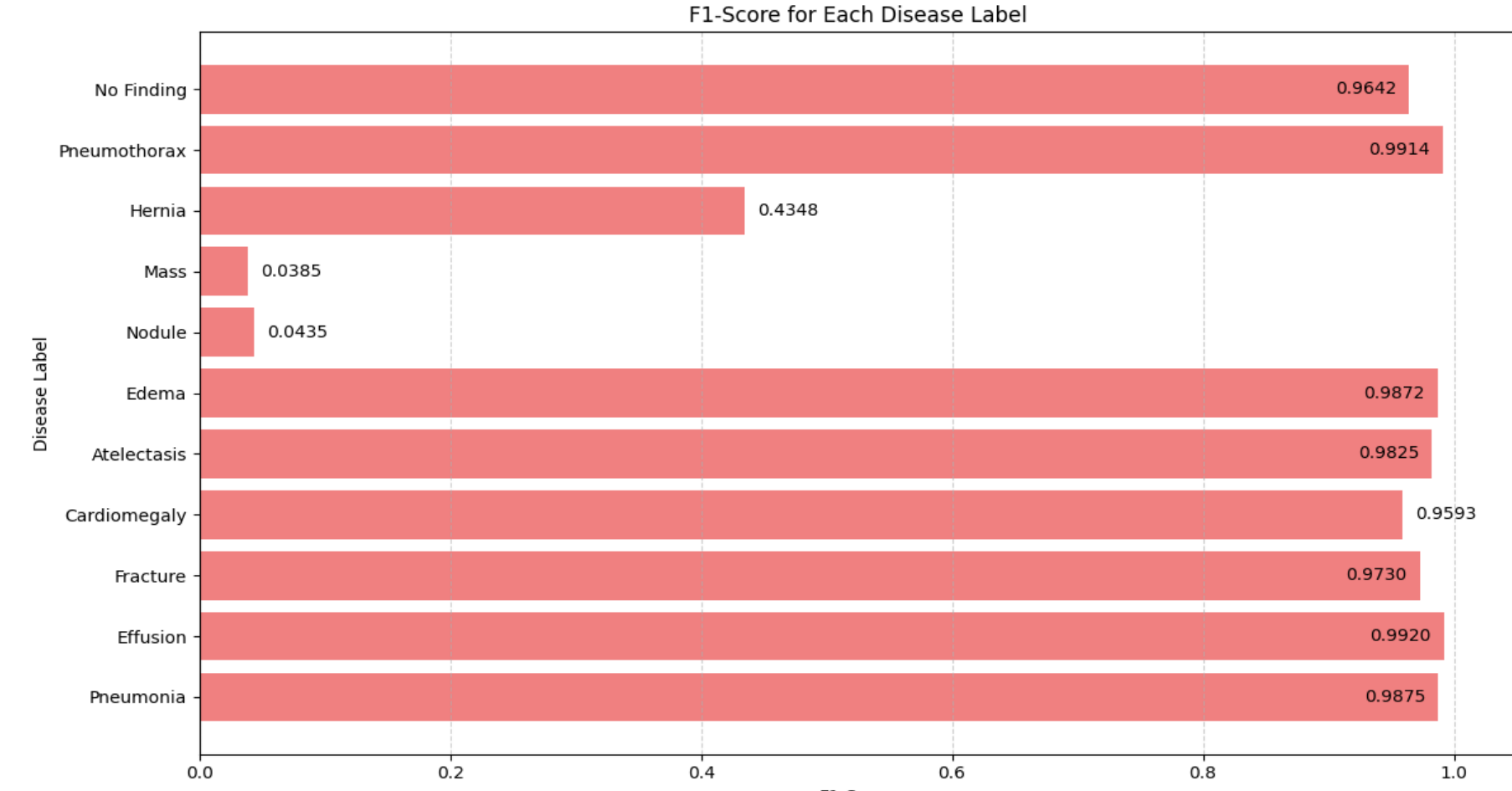**Fig. 8:** Regression calibration analysis.

Fig. 9: Pathological performance.

CONCLUSIONS

Integration: MTN successfully fuses visual/textual data for comprehensive CXR analysis.

Clinical Impact: Provides interpretable outputs for actionable triage ranking, serving as a scalable decision support tool.

FUTURE DIRECTIONS

1. **Scale Training:** Expand to the full MIMIC-CXR dataset to improve detection of rare pathologies.
2. **External Validation:** Test on independent hospital datasets to ensure cross-institutional reliability.
3. **Temporal Analysis:** Integrate longitudinal data to monitor disease progression and stability.
4. **Clinical Pilot:** Deploy in live workflows to measure real-world reduction in triage delays.

REFERENCES

1. B. Oltu, S. Güney, S. E. Yuksel, and B. Dengiz. Automated Classification of Chest X-Rays: A Deep Learning Approach with Attention Mechanisms. BMC Medical Imaging, 25:71, 2025.
2. J. Xiao, S. Li, T. Lin, J. Zhu, X. Yuan, D. D. Feng, and B. Sheng. Multi-Label Chest X-Ray Image Classification with Single Positive Labels. IEEE Transactions on Medical Imaging, 2024.
3. X. Liu, Y. Liu, H. Chen, et al. MDFormer: Transformer-Based Multimodal Fusion for Robust Chest Disease Diagnosis. Electronics, 14(10):1926, 2025.
4. Y. Li, H. Wang, and Y. Luo. A Comparison of Pretrained Vision-and-Language Models for Multimodal Representation Learning Across Medical Images and Reports. arXiv:2009.01523, 2020.
5. G. Jacenków, A. Q. O'Neil, and S. A. Tsaftaris. Indication as Prior Knowledge for Multimodal Disease Classification in Chest Radiographs with Transformers. arXiv:2202.06076, 2022.
6. M. Lin, Z. Wang, Y. Zhou, et al. An Empirical Study of Using Radiology Reports and Images to Improve ICU-Mortality Prediction. arXiv:2307.07513, 2023.