Cross-Dataset Generalization of AI Action Recognition Models for Combat Casualty Care in Unseen Environments

Yupeng Zhuo[1], Brian VanVoorst[2], Oanh Tran[3], Andrew W. Kirkpatrick[4], Kyle Couperus[3], Jessica McKee[4], Nick Walczak[2], Lee Tarlin[2], Christopher Gilleo[2], Charles Meissner[2], Juan P. Wachs[1]

[1]Edwardson School of Industrial Engineering, Purdue University, West Lafayette, Indiana, USA; [2]RTX BBN Technologies, Cambridge, Massachusetts, USA; [3]Exonicus, Inc., Seattle, Washington, USA; [4]University of Calgary, Calgary, Alberta, Canada.

Email: jpwachs@purdue.edu

Motivation

- **Medic Cognitive Overhead:** Combat medics operate under severe physical stress, hostile noise, and critical time pressure in far-forward care (Roles 1 & 2). Systems requiring manual context-switching across separate single-task algorithms add intolerable cognitive burden.
- **Data Curation Bottleneck:** Curating massive, fully annotated expert video datasets for every conceivable combat casualty procedure is logistically, ethically, and computationally prohibitive in battlefield settings.
- **In-Distribution Vulnerability:** Conventional clinical AI relies on static in-distribution training data, causing catastrophic failure when deployed to unpredictable operational environments or novel emergency procedures.

- **Across MViTv2, Qwen2.5VL, Qwen3VL, and Qwen3.5, training exclusively on the small Trauma THOMPSON dataset enabled recognition of medically similar actions in an unseen EgoMAGIC environment.**

Research Paradigm & Evaluation Strategy

- **Core Hypothesis**: Emergency medical care consists of universal physical action primitives (e.g., limb manipulation, strap tensioning, tube insertion). AI models trained on core primitives can recognize novel complex procedures based on biomechanical similarity.
- **Semantic Merging & Metric:** Implemented semantic class merging via all-MiniLM-L6-v2 text embeddings (Cosine similarity threshold = 0.95). Evaluated via Transfer Efficiency, normalizing zero-shot transfer against structural dataset mismatch.
- **Transfer Efficiency:** Normalizes zero-shot transfer performance by isolating the AI's ability to recognize actions it has seen before, accounting for structural mismatch between the datasets. It is defined as:

$$\text{Transfer Efficiency} = \frac{\text{Correct Actions Recognized in EgoMAGIC}}{\text{Total Overlapping Actions (EgoMAGIC} \cap \text{Trauma THOMPSON)}}$$

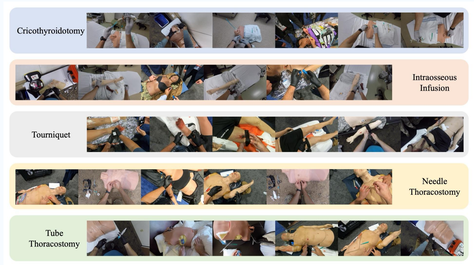
Cross Dataset Generalization Framework

- **Input:** The AI receives sequential egocentric video frames captured from a medic's viewpoint.
- **Processing:** vision-only model (MViTv2) and multimodal vision-language models (Qwen2.5VL, Qwen3VL, Qwen3.5)
- **Output:** The model predicts the specific medical action primitive taking place (e.g., strap tensioning, limb stabilization, mask placement).
- **Cross-Dataset Framework:** Models were trained on the small dataset and tested zero-shot on the massive and unseen dataset

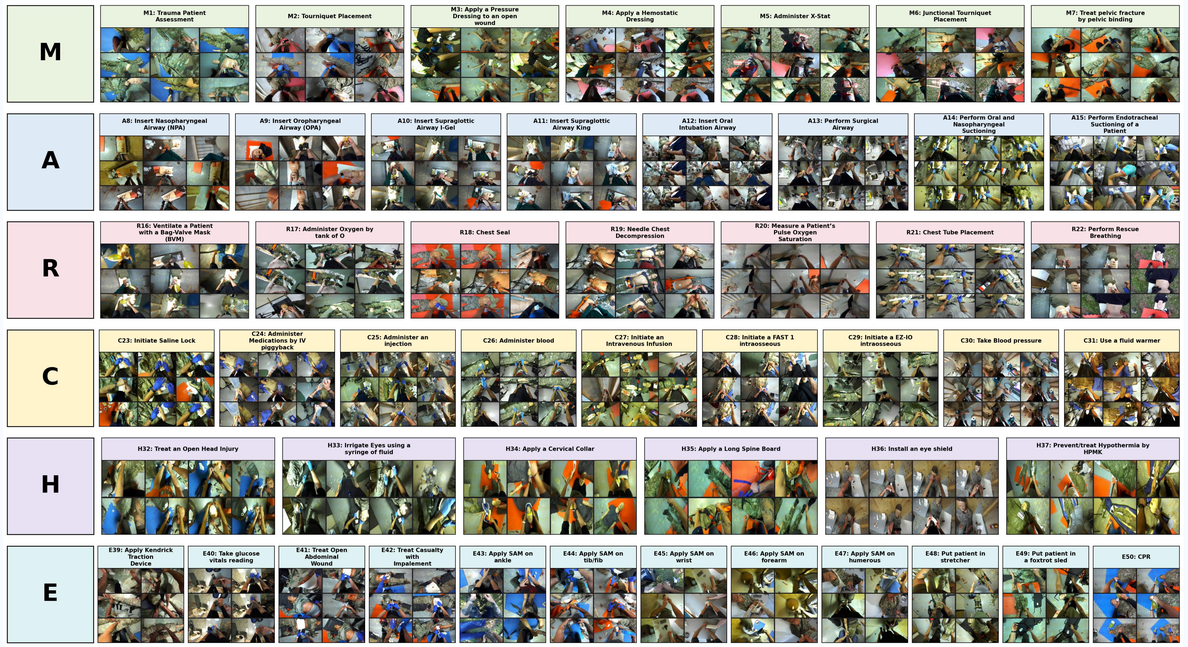
TRAIN
Trauma THOMPSON
5 procedures
220 videos

LEARN
Shared biomechanical primitives: compress, insert, transfer, stabilize

TEST
Unseen EgoMAGIC
50 procedures
2,890 videos



Small Scale Dataset
Trauma THOMPSON Dataset [1]



Large Scale Dataset
EgoMAGIC Dataset [2]

Cross-dataset Generalization Performance Comparison

Code & Category	Unseen Target Procedure Name	MViTv2 Efficiency	Qwen2.5-VL Efficiency	Qwen3.5 Efficiency	Qwen3-VL Efficiency	Underlying Biomechanical Feature
E48 / E49 (Transport)	Patient Stretcher & Sled Transport	**52.63%**	10.83%	10.11%	14.80%	Whole-body repositioning & lifting kinetics
E39 (Orthopedic)	Kendrick Traction Device	**49.41%**	8.16%	1.77%	11.70%	Limb stabilization & strap tensioning
R22 (Airway/Respir.)	Perform Rescue Breathing	0.00%	0.00%	4.30%	**60.87%**	Mask placement & chest rise cadence
E42 (Trauma Care)	Treat Casualty with Impalement	0.00%	0.00%	**38.46%**	15.38%	Wounding site packing & dressing application
M2 / M6 (Hemorrhage)	Tourniquet & Junctional Tourniquet	30.04%	5.00%	9.74%	**36.28%**	Circumferential limb constriction & torque
H33 / H36 (Ocular Care)	Eye Irrigation & Eye Shield Installation	0.00%	19.29%	25.00%	**32.14%**	Fine motor facial manipulation & syringe wash

Key Results & Core Findings

- **Biomechanical Similarity Drive:** Cross-dataset generalization efficiency is governed by kinetic and movement similarity (e.g. limb manipulation, strap tensioning, tube insertion) between source and target procedures.
- **Architectural Strengths:** MViTv2 vision backbone excels at physical movement tasks (Stretcher Transport: 52.6%, Traction Device: 49.4%). Qwen3-VL/Qwen3.5 VLMs excel at semantic reasoning tasks (Rescue Breathing: 60.9%, Eye Irrigation: 32.1%).
- **Zero-Shot Novel Procedure Success:** Models accurately identified action steps across 22+ novel, unseen medical procedures without receiving any training frames from those target scenarios.

60.87% PEAK TRANSFER EFFICIENCY
Achieved zero-shot by Qwen3-VL on unseen Rescue Breathing (R22) procedures, demonstrating semantic action recognition driven by biomechanical similarity to airway interventions.

49,490 VIDEO CLIPS EVALUATED
Rigorous zero-shot evaluation across 50 complex unseen combat casualty procedures in the EgoMAGIC dataset, validating robust cross-dataset generalization from only 5 training procedures.

Operational Impact

- **Eliminates Data Bottleneck:** Demonstrates field medical AI assistants do not require exhaustive video collection for every trauma scenario, shortening deployment timelines.
- **Zero Cognitive Switching:** Shows a single, unified AI assistant operating continuously in the background, freeing medics from manual model selection during emergency care.
- **Unseen Scenario Resilience:** Ensures decision support systems maintain efficacy during unexpected operational casualties or novel combat care procedures.

Future Work

- **Wearable Edge Deployment:** Port compressed vision-language models to lightweight head-mounted displays (e.g. HoloLens / smart glasses) for far-forward field testing.
- **Dynamic Prompting Adaptation:** Integrate real-time text/visual prompting to rapidly adapt the AI assistant to emerging and rare tactical medical procedures.

References

[1] Y. Zhuo et al., "An Egocentric Life-Saving Interventional Procedure Dataset of Actions, Medical Questions, Maneuvers and Tools," Sci Data, vol. 13, no. 1, p. 51, Dec. 2025, doi: 10.1038/s41597-025-06365-y.
[2] B. VanVoorst et al., "EgoMAGIC- An Egocentric Video Field Medicine Dataset for Training Perception Algorithms," 2026, arXiv. doi: 10.48550/ARXIV.2604.22036.

Acknowledgement

This work was supported by the US Army Medical Research and Development Command under Contract No. W81XWH21C0119 and by the National Artificial Intelligence Research Resource Pilot under Grant NAIRR250333.

Disclaimer

The views expressed are those of the author(s) and do not reflect the official policy of the Department of the Army, the Department of Defense or the U.S. Government. The investigators have adhered to the policies for protection of human subjects as prescribed in 45 CFR 46.