AWARD NUMBER: W81XWH-17-C-0252

TITLE: Automatic Sensing for Clinical Documentation (ASCD)

PRINCIPAL INVESTIGATOR: Daniel V. Fabbri, PhD.

CONTRACTING ORGANIZATION: Vanderbilt University Medical Center, Nashville, TN

REPORT DATE: April 2021

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release; Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE
April 2021

2. REPORT TYPE
Final

3. DATES COVERED
30Sep2017– 29Dec2020

4. TITLE AND SUBTITLE
Automatic Sensing for Clinical Documentation (ASCD)

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-17-C-0252

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Daniel Fabbri, PhD; Laurie Novak, PhD, MHSA; Julie Adams, PhD;
Robert Bodenheimer, PhD; Joseph Coco; Christopher Simpson; Jamison Heard;
Devin Rickard; Deirdre Scully; Candace McNaughton; Devin Rickard; Deirdre Scully

E-Mail:

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Vanderbilt University Medical Center
Department of Biomedical Informatics
2525 West End Avenue
Nashville, TN 37203-8556

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Current communication methods and clinical documentation during from the point-of injury through the phases of care in the military's operational environment continues to be incomplete, inaccurate, and hinders receiving facilities ability to rapidly gain situational awareness of the patients. This project is developing a novel hands-free system that automatically detects the motion signatures associated with key clinical tasks to generate and transmit a care record in real-time that ensures better, more consistent, and clear communication among care teams.

15. SUBJECT TERMS
Biomedical informatics; medical clinical documentation; accelerometers; video; hands free; Tactical Combat Casualty Care TCCC; Trauma; activity detection recognition; injury heatmap

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	100	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

1. INTRODUCTION:

Clinical documentation during both the point-of injury and en route phases of care in theater and operational environments continues to be incomplete, inaccurate, and detrimental to the goal of ensuring that receiving facilities can rapidly gain situational awareness of the patients moving through the system. Current communication methods between the point of injury and receiving facilities rely on verbal and written communication. These methods are vulnerable to rapid changes in clinical status and human cognitive biases in data collection, processing, and sharing. And, multiple handoffs further complicate the process and increase risk of errors and miscommunication.

This project developed a novel hands-free clinical documentation system for use in the operational environment that leverages a combination of off-the-shelf sensors, accelerometers, and cameras to build a software system that automatically detects the motion signatures associated with key clinical tasks and generates an abbreviated care record, which can be transmitted upstream in real-time. The project's intent is to ensure better, more consistent, and clear communication among care teams to overcome human misperceptions and error, provide high fidelity and reliable data, and allow communication across multiple patients and providers at the same time.

2. KEYWORDS:

Biomedical informatics; medical clinical documentation; accelerometers; video; hands free; Tactical Combat Casualty Care TCCC; Trauma; activity detection recognition; injury heatmap

3. ACCOMPLISHMENTS:

Project Objectives and Major Aims. The project's hypothesis is that the automatic identification, documentation, and communication of key clinical concepts (i.e. injury patterns or clinical interventions) that occur during the initial phase of care (i.e. point-of injury and en-route care) will satisfy the information needs of upstream care providers and facilitate better care coordination and resource utilization. The project's overall objective is to create a novel hands-free system using wearable technology and cameras that can improve care by automatically sensing, documenting, and transmitting clinical events with little or no end-user input.

This objective is supported by the specific aim to design and implement a clinical activity detection prototype system using accelerometers, electromyography (EMG) and video. The project includes two milestones: 1) the development of prototype software and 2) drafting of publications on the problem and methods. The project's aim, and the milestones, are underpinned by the following five major tasks. (See Table 1 for the Aims, Major Tasks, and Milestones, and Appendix A for the full Project Scope of Work and Breakdown Schedule).

1. The first major task that the project is pursuing is the development of clinical activity detection algorithms utilizing accelerometer data. To accomplish this, the

project 1) designed and evaluated methods to extract accelerometer data from devices in real-time or near real-time; 2) designed and implemented basic activity detection algorithms using captured accelerometer data; and 3) aggregated the accelerometer data into a centralized physician dashboard.

2. Another major task is developing clinical activity detection algorithms that utilizes image (video) data. The accomplish this, the project: 1) developed systems to capture video data using cameras; 2) designed and implemented a basic clinical activity detection system using image (video) data; and aggregated the data (image) data into a centralized physician dashboard.

3. The next major task combined the accelerometer and image (video) data to develop clinical activity detection algorithms, which included: 1) developing models to correlate accelerometer and image data for activity detection; and 2) evaluating combined activity detection algorithms on simulation lab data.

4. The fourth major task was designing and implementing high-level clinical activity features that include: 1) designing and implementing an injury heatmap visualization; 2) designing and implementing a risk score derived from accelerometer data; and 3) designing and developing 'quick' data entry systems that prints results on top of a TCCC card.

5. The last major task was conducting focus groups and field data collection. This has been broken down broadly into task: 1) focus group of medics, surgeons, emergency department physicians, and military staff; and 2) field data collection deploying the prototype systems with Nashville Fire Department paramedics.

AIM(s) & MILESTONE(s)	**DATES**		**WORK %**		**STATUS**
Description	Start	End	Target	Complete	
Specific Aim 1: Design & implement a clinical activity detection prototype system	01OCT17	31DEC19	100%	100%	On-Track
Milestone #1: Prepare publication on problem overview, and detection algorithms	01DEC18	30SEP19	100%	100%	On-Track
Milestone #2: Package prototype software for sharing and distribution	01MAY19	29DEC20	100%	100%	On-Track
Milestone #3: Complete and submit final report	01OCT19	29DEC20	100%	100%	On-Track

MAJOR TASKS		**DATES**		**WORK %**		**STATUS**
#	Description	Start	End	Target	Complete	
1	Develop clinical activity detection algorithms utilizing accelerometer data	01OCT17	30NOV18	100%	100%	On-Track
2	Develop clinical activity detection algorithms that utilizes image (video) data	01OCT17	30NOV18	100%	100%	On-Track
3	Develop clinical activity detection algorithms that combines accelerometer & image data	01JUN18	29DEC20	100%	100%	On-Track
4	Design and implement high-level clinical activity features	01MAY19	30SEP19	100%	100%	On-Track
5	Focus group and field data collection of developed prototype systems	01NOV17	30SEP19	100%	95%	On-Track

** On-Track = +/-5; Ahead >+5%; Behind >-%5*

Table 1: Aims, Major Tasks, and Milestones

Accomplishment towards achieving the Aims.

The team achieved several notable accomplishments towards completing the project's aim. This includes:

1. Identified key domains and design insights from focus groups.
2. Affirmation of the premise that there is a need for improved documentation between medics and hospital personnel.
3. Developed an accelerometer and EMG data capture system using Myo devices.
4. Developed an accelerometer and motion data capture system using Apple Watches.
5. Developed a Documentation Dashboard served from a webserver.
6. Developed a procedure prediction system using Myo and Apple Watch data.
7. Developed a procedure prediction system using video data.
8. Developed a procedure prediction system using motion and video data.
9. Designed and implemented an injury heatmap visualization.
10. Deployed the prototype system to capture data in the field with Nashville Fire.
11. Developed an auto-generated Tactical Combat Casualty Care card.
12. Published associated results in conference and journal proceedings.

These accomplishments, and the preparation and process that led to them, are discussed in more detail below.

Domain Identification and Development of Design Insights.

The project conducted three focus groups comprising 13 participants, which was approved by the Vanderbilt University Institutional Review Board. Two of the focus groups included pre-hospital personnel and one was conducted with hospital personnel. The focus groups were audio recorded and transcribed for analysis. The transcriptions were analyzed using a qualitative data analysis tool (Dedoose™). Data, which was coded using a taxonomy, resulted in a key takeaways document that summarized the design insights from the data across six key domains.

The goals were to gather information from healthcare providers with trauma experience to identify gaps in current handoff procedures and understand current documentation practices. The sessions focused on the process of communicating during transitions of care from EMT to hospital, including elicitation of actual experiences in the combat environment when possible. Based on the information shared in the session, probing questions were added to better understand the physical actions involved in transporting patients from the field/scene to the hospital including the implications of incorporating wearable technologies, cameras, and other devices into the process. The focus group guide is included in Appendix B.

The study created a list of key findings within the key domains that were generated from the focus groups (see Appendix C for the complete list). From those, the project identified several major findings that continue to provide significant insight into the system's design.

For instance, in the domain of Vital Signs and Demographics, the key design insight from the focus groups was the need to: 1) corroborate the vital signs, and 2) capture as many of the important, routinely given vitals and demographics. Interestingly, hospital personnel asked for two measurements: the current value and the most extreme.

For the domains of Medication Administration and Documentation, the major findings were to: explore how to automatically capture the administration of medicine, and corroborate the medicines administered. In addition, it was identified that EMTs use ad hoc methods to informally document their procedures (e.g., writing on tape on their leg), which the administrations are not formally documented until after transport. This showed the need to: 1) review the procedures EMTs document to better understand which procedures merit automatic identification, and 2) explore the possibility of transmitting the captured information to the EMT or hospital in real or near-real time.

For the Transport domain, the groups major finding was the necessity for: 1) the corroboration of the procedures done in transport and their sequence (sequence of procedures matters so procedures need to include times), and 2) that real time information provided to the receiving hospital will allow the efficient allocation of resources and triage.

The Communication domain illustrated that the capturing, processing, and sharing the real time information from EMTs to the receiving hospital will reduce miscommunication and allow focus on care to the patient(s).

For the last domain, Incorporating (New) Technology, the study found that most people are amenable to wearable, body tracking, machine learning technologies, but additional thought is needed on how to best overcome the cultural "trust" gap with new technology. The study also found for the need to explore the ergonomics of wearing the devices with required uniforms and within the physical space limitations of transport vehicles.

<u>Identification of accurate medical information.</u>

The project identified critical elements of information that must be captured and communicated by observing previously recorded handoffs between pre-hospital and hospital personnel for trauma patients. To accomplish this, **the team reviewed 50 Level I trauma resuscitation videos that are regularly taped for quality improvement purposes**. These videos capture the pre-brief (in which trauma team members from the emergency department and trauma team review known facts about the case and discuss a plan of action), hand-off, and management while the patient is in the trauma bay.

To develop data capture form and coding scheme for the reviews, three reviewers reviewed five videos. After all the videos had been reviewed, the reviewers met to discuss the results and any discrepancies between reviewers. The reviewers then came to a consensus about the types of information transferred from pre-hospital to hospital personnel and developed a codebook to be applied by a single observer.

A total of 50 trauma videos were reviewed, and information from the associated handoffs were recorded. These videos were reviewed by a single observer, who was trained as a nurse and has extensive experience with trauma and the videos reviewed. Data was collected on an observation form that was created specifically for that purpose. The observational form was edited after the initial five video assessment activity to reflect the information that was most pertinent in the video interactions. After completion of the 50 reviews, the results from the observation forms were entered into a REDCap database for further analysis and tabulation.

A major finding from reviewing the videos illustrated that important information, needed to provide optimal care, is not always effectively conveyed during the handoff. For example, many clarifying questions were asked during handoffs. Clarifying questions were found in 40 of the 50 videos from the hospital staff during the handoff from pre-hospital personnel. The presence of clarifying questions during handoff(s) reflects a need for the information to be relayed more accurately to meet the needs of trauma team personnel. Moreover, the video review confirmed the set of procedures the team should attempt to automatically detect such as intubation, IV, and bagging, among others. See Appendix D for all findings from the video observations.

The pre-hospital and hospital teams have different priorities and/or capabilities in the performance of their roles in their respective environments. Pre-hospital teams need to get the patient in the vehicle, perform needed procedures during transport to stabilize the patient, and deliver the patient to more capable facility, which usually provides surgical intervention. Meanwhile, the receiving trauma team wants to be able to appropriately allocate resources in advance, based on case complexity. These differences seem to result in an inadvertent conflict about the priority of recording specific times of medication administration and/or performance and sequence of procedures during transport, and the uninterrupted care of the patient.

Data from the observations supports the findings from the three focus groups that more accurate information is needed at the time of handoff, specifically regarding time and sequences of procedures and/or medications. The hospital focus group detailed that the most important information needed by the trauma team involved time, specifically regarding the sequence of procedures performed during transport. Specifically, doctors care about how often and what a medic is doing during transport. In summary, the results of the focus groups and video observations have illustrated the need for more accurate recordkeeping, specifically temporal aspects, to enhance the handoff from pre-hospital to hospital teams.

Development of a Myo reporting device.

The Myo Gesture Control Armband is a wearable sensor device consisting of electromyography (EMG) and inertial measurement unit (IMU), as shown in Figure 1. It has been used in medical environments and in research projects for activity detection.

Figure 1: Myo Gesture Control Armband

While Myo devices have configurable, integrated hardware for gesture prediction of five to six pre-defined gestures, it does not support activity prediction. Myo does not support simple storage and egress of raw sensor data, but it has released a feature-rich software developer's kit (SDK). We have leveraged this SDK, in conjunction with the Myo Connect application, to create a basic desktop multi-Myo recording application called Myo Egress. With Myo Egress acting as a real-time EMG and IMU sensor storage system, we can use state-of-the-art machine learning to predict medic activity from their movements and generate documentation.

Myo Egress is a python 3 application which runs on MacOS and Windows. It's capable of recording and transmitting data in JSON or CSV formats. The IMU data consists of gyroscope: x-rotation, y-rotation, z-rotation; acceleration: x-acceleration, y-acceleration, z-acceleration; and orientation: w-quaternion, x-quaternion, y-quaternion, z-quaternion all at 60 Hz. Eight EMG sensors also record electrical impulses translated into an 8-bit signed integer representing electrical intensity at 200 Hz. Additionally, higher-level data such as pose prediction, state data, such as lock, and metadata such as bluetooth signal strength are stored at 200 Hz as well. In addition to Myo Egress recording data locally to a desktop computer, it can simultaneously record data in bursts of 1 Hz to multiple web services, such as the Handsfree Documentation web service designed to analyzing sensor fusion data of IMU, EMG, and video.

The Myo Armband has some basic feedback mechanisms primarily consisting of short and long vibrations. Currently, we have chosen not to use these features as a passive recording environment is preferred. However, the Myo Connect application vibrates Myo devices when they desynchronize, or when a state is entered in which data quality is degraded until the device is recalibrated. We are in the process of bypassing the Myo Connect application so all vibrations from the Myo Armband can be halted. In the lab simulations using medical mannequins, data have been collected that are complete and consistently high quality. Neither vibrations nor wearing the devices have interfered with medics' work. In the future, we may also write Myo Egress to run on a smartphone so a laptop does not need to be present to record medic data.

Development of an Apple Watch reporting device.

The Apple Watch is a wearable iPhone-synced, sensor capable of recording inertial measurement unit (IMU), as shown in Figure 2. It also has been used in medical environments and in research projects for activity detection.

MacOS enables any user to design and build their own application for the Apple Watch and iPhone. We have created such an application to record accelerometer data on the Apple Watch over any length of time, with the simple touch of a start and end button. The data are streamed to a storage system in batches. With this accelerometer data we can use state-of-the-art machine learning to predict medic activity from their movements and generate documentation.

Another advantage of the Apple Watch and synced iPhone is that the team can track the geolocation of medics wearing the device. This allows the team to receive a notification when the wearer is within a certain mileage range of a hospital.

Figure 2: Apple Watch and iPhone 6 Synced

The Apple Watch application allows the user to denote whether the watch is worn on the right or left wrist. When started, it records continuous accelerometer data at 60 Hz. Currently each Apple Watch must be synced to its own iPhone for the application to run. In addition, the iPhones must be connected to WIFI or have cellular data access for the accelerometer data to be sent to our data storage system.

The IMU accelerometer data consists of gyroscope: x-rotation, y-rotation, z-rotation; acceleration: x-acceleration, y-acceleration, z-acceleration; and orientation: yaw, pitch, roll all at 60 Hz. The Apple Watch application sends the accelerometer data automatically in real-time to our data storage system web application. This web application, the Handsfree Documentation web service, was designed to analyze sensor fusion data of IMU and EMG between the Myo and Apple Watch, and video recordings.

The current system being tested consists of two Apple Watch 3s synced to two iPhone 6s. Newer versions of both the Apple Watch and iPhone can also be used, but the Apple Watch 3 and iPhone 6 are the necessary minimum versions to run the current application. The smallest available iPhone memory option is also sufficient to host the Apple Watch and application.

Development of a Handsfree Documentation Dashboard

In order to fuse the ten streams of data coming from each medic, the Handsfree Documentation Dashboard was created as a real-time web service. Graphs of Apple Watch acceleration, Myo Armband acceleration, and EMGs are synced with 4 video feeds (see Figure 3). This dashboard helps to visually analyze the data collected and iteratively improves the system's design.

The Handsfree Dashboard is a secure, HIPAA-complaint central storage repository for sensor and video data and metadata. Metadata syncs our data streams which allows for human analysis, validation, and pattern recognition of our data. Comparing participants side-by-side is possible by selecting each participant in an experiment in separate windows and selecting the events to compare, such as administering medication through IV or chest-tube decompression. Additionally, IMU sensor data can be validated against movements seen in video and also against the two IMU sensors per arm on each participant.

Comparing IMU sensors is more difficult than we initially anticipated because Apple Watch sensors subtract gravity from their output, but Myo Armband sensors do not. Once gravity has been added back to the Apple Watch, and gravity removed from the Myo Armband, those two streams will be used in the dashboard in conjunction with the video to determine if the Myo Armband IMUs, Apple Watch IMUs, or both IMUs are best for activity prediction. (See Appendix F for EMG and Accelerometer Graphs)

Figure 3: Synced four camera angles of Handsfree Dashboard

Accelerometer data analysis and procedure prediction.

The collected accelerometer and electromyography (EMG) data were first analyzed qualitatively to determine patterns that persist across participants and to drive the feature extraction process. The IMU and EMG data for each medical procedure were plotted to determine patterns or state-changes that appear in each instance of the medical procedure. A plot of the acceleration data for CPR captured by the Myo device is provided in Figure 4**.**

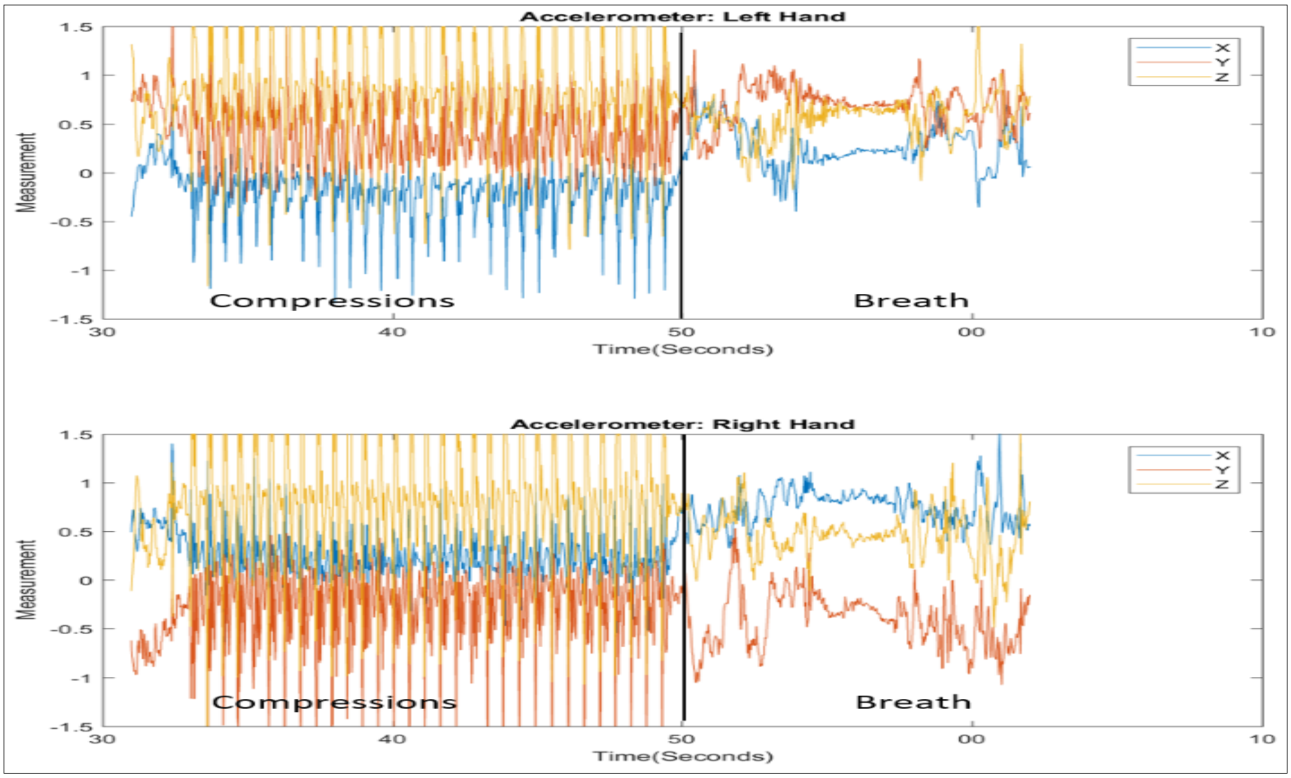


Figure 4: Myo Acceleration Data for CPR

A sinusoidal signal occurred when the participant was performing chest compressions is seen in the left hand of the figure. An abrupt change occurred in the acceleration data when the participant gave the patient two breaths, which indicates a state-change. Not all medical procedures elicited such clear patterns in the IMU data, such as using a bag-valve mask to ventilate a patient, but sometimes a pattern occurs in the EMG data. A plot of the EMG data captured by the Myo device on the participant's right hand, which was used to squeeze the bag, is provided in Figure 6. The IMU data for bagging remained stationary, while high amplitude periods occurred in the EMG data. Each high amplitude period represents the participant squeezing the bag, where the EMG data in Figure 5 shows a total of seven squeezes. The number of squeezes was verified by examining the video data.

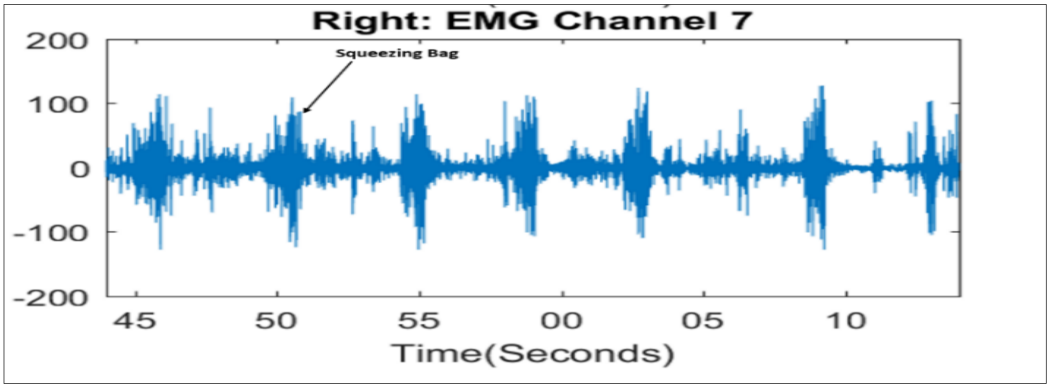


Figure 5 Myo EMG Data

Certain medical procedures did not have clear patterns, such as inserting an oral airway, which may be attributed to the quick movements needed to do the procedure. Additional training data was determined to be needed for these medical procedures (and was collected) for a machine-learning algorithm to achieve high classification accuracy.

Descriptive statistics were calculated for each procedure and participant observed in the simulation lab to examine statistical differences within a participant, and between procedures and participants. The descriptive statistics for CPR by participant for the Myo's Right Hand IMU data is provided in Table 2. Although, CPR produces a sinusoidal pattern in the IMU data, the descriptive statistics show that there are differences between participants for the Myo's IMU data. The differences are attributed to sensor drift, as the Myo device does not subtract gravity from the IMU data. The Myo's EMG data does produce similar descriptive statistics for CPR for each participant, which is attributed to the EMG signal being a zero-mean signal. This analysis illustrates that other features need to be extracted from the IMU and EMG data for a classifier to achieve high performance.

Frequency-based features, such as entropy, were shown to produce significant differences between procedures most of the time, while not producing significant differences between instances of the same procedure. The results of this statistical analysis are not provided to keep the report concise, due to the breadth of the procedures covered. The analysis provided the foundation for the feature extraction process and selection of features for machine-learning.

Participant	Acc_X	Acc_Y	Acc_Z	Roll	Pitch	Yaw
P1	-0.07 (0.31)	0.78 (0.54)	-0.51 (0.41)	-1.67 (0.3)	-0.95 (0.35)	-1.94 (1.31)
P2	0.24 (0.26)	-0.73 (0.3)	0.6 (0.43)	-1.25 (0.13)	-0.65 (0.29)	-0.27 (0.48)
P3	0.43 (0.23)	-0.44 (0.39)	0.72 (0.38)	-0.73 (0.37)	-0.82 (0.22)	1.65 (0.38)
P4	-0.51 (0.48)	0.79 (0.60)	0.04 (0.29)	-1.43 (0.23)	-0.94 (0.26)	-2.37 (0.91)

Table 2: Participant Table

Time-based and frequency-based features were extracted from the Myo device's IMU and EMG data and have been used in prior wearable-sensor activity recognition algorithms.[1] The average, max, and standard deviation of each IMU and EMG signal was extracted over the entirety of each medical procedure. Then, each signal was transformed into the frequency domain to calculate the DC power, total power, and entropy of the signal.

Design and implementation of injury heatmap visualization.

Video data collected during over 15 hours in the simulation lab were processed through OpenPose. The resulting files represent the position of each person and their body parts in every single frame of the video with sets of X and Y coordinates and their confidence intervals (CI). The confidence interval describes how sure Open Pose is for that specific body party location. Open Pose processes the following body parts for each person present in every frame of the video, as shown in Figure 1: Nose (0), Neck (1), Right Shoulder (2), Right Elbow (3), Right Wrist (4), Left Shoulder (5), Left Elbow (6), Left Wrist (7), Right Hip (8), Right Knee (9), Right Ankle (10), Left Hip (11), Left Knee (12), Left Ankle (13), Right Eye (14), Left Eye (15), Right Ear (16), Left Ear (17), as shown in Figure 6.

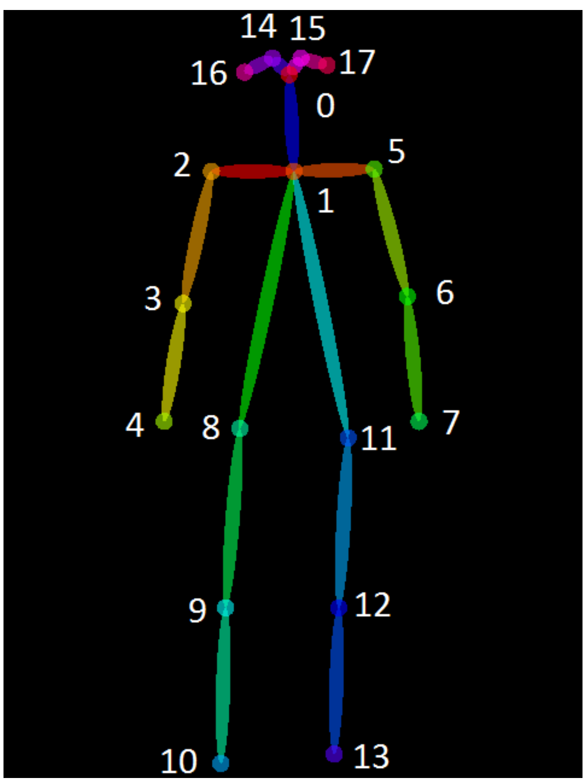


Figure 6 Key points tracked over the body

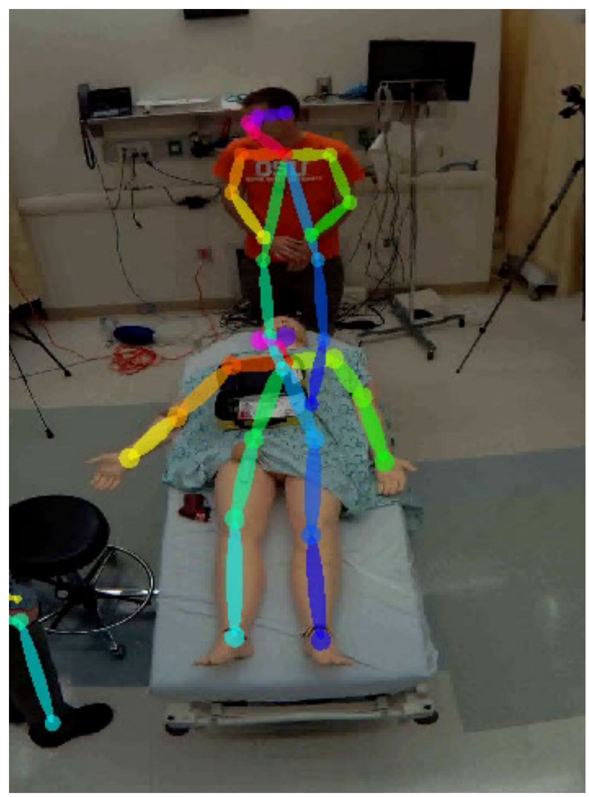


Figure 7: Open Pose Single Frame Output Example

[1] Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., & Amirat, Y. (2015). Physical human activity recognition using wearable sensors. *Sensors*, *15*(12), 31314-31338.

From the Open Pose output, the next step is identifying the medic and the patient in the frame. The patient is identified as the person closest to the center of the image frame. The medic is then identified as the second closest person to the center of the image frame, and all other persons are ignored. Once the medic and patient are identified based on the positional assumptions, the medic's left and right-hand locations, and all of the patient's body part locations are extracted. Figure 7 shows how the Open Pose human figure is automatically applied to each person present in every frame of the video.

Next heatmaps are created based on where the medic and their hands are located throughout the procedure. To generate a heatmap, the system generates a Gaussian field around each medic hand position per frame and summed over all frames. By summing, the heatmap captures the most frequently occurring positions of the hands over the patient's body per procedure.

Figure 8 shows a heatmaps that was generated using the IV procedures. The background image represents the patient's body, and the colors represent the position of the EMT's hands over the patient's body in patient space. The yellow color represents the areas above and around the patient where the EMT hands are located most often. Visually these heatmaps indicate that we can identify the body part, which is being worked on, which will help in determining which procedure is being performed.

Specifically, the heatmaps were registered to align the patient to the TCCC standard card. This alignment allows for comparison of different cases. Registering patients to the standard TCCC card requires managing the slight translation of the patient's limbs to the limbs on the TCCC card.

This work presents a method to visualize heatmaps that allows the inspection of a given set of procedures. Of note, the work was recently accepted for publication at the conference SPIE.[2]

Since the project intends to use this in conjunction with activity data gathered from other devices, this work shows a first step in how computer vision and machine learning can be used to help further identify the procedure being performed.

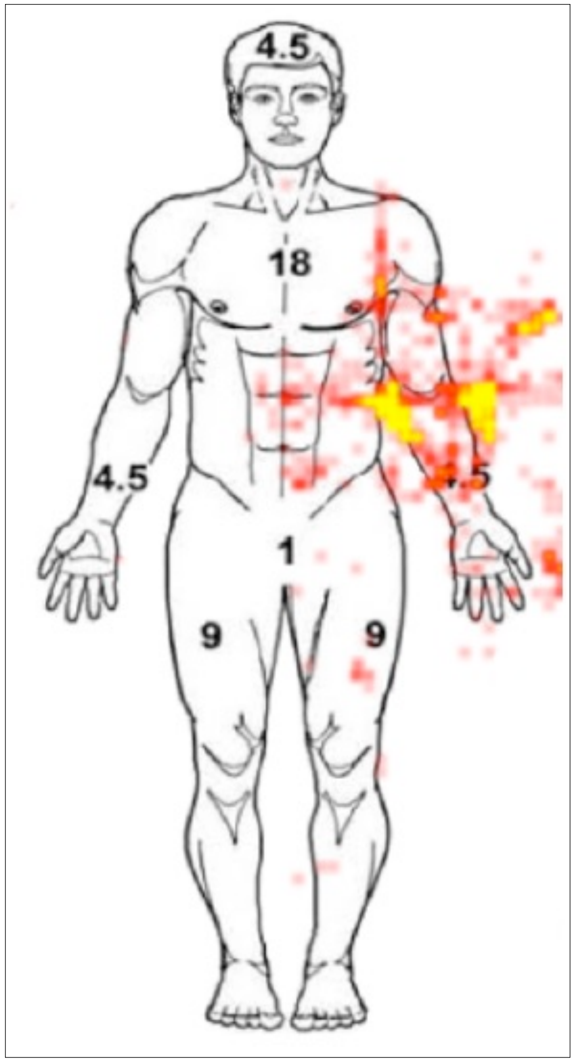


Figure 8 Heatmap of medic hand position over body for IV

Emergency Severity Index (ESI) Classification.

Civilian and military hospitals are faced with the challenge of providing efficient care with limited, sometimes scarce, resources. Having accurate, timely information regarding patients being transported greatly assists in efficiently and safely allocating scarce resources including blood products, clinical expertise and time, and facilities such as operating theaters. The Automated Sensing Clinical Documentation System (ASCD), which monitors medics using inertial measurement units (IMU) and electromyography (EMG) sensors, can be modified to predict patient emergency severity index (ESI, which is a rapid triaging system based on illness severity and anticipated resource utilization) and allow for more efficient matching of predicted resource utilization with available resources. Gathering enough real-world medic measurements and ESI scores to assess predictive relationships with sufficient statistical power to requires dozens of ambulance ride-along, but our preliminary feasibility testing has indicates that the mean of energy expenditure is strongly inversely correlated with ESI triage score on ED arrival (rho -0.8, 95% -0.88 to –0.22).

Partnering with the Nashville Fire Department, a research observer rode in ambulances with NFD paramedics equipped with our data sensors with 64 patients on 35 shifts. For 22 pre-hospital transportation events, we were able to pair ESI scores with medic IMU, medic EMG, vehicle IMU, and manual event logs. Event logs were

[2] Paris, R. A., Sullivan, P., Heard, J., Scully, D., McNaughton, C., Ehrenfeld, J.M., Adams, J.A, Coco, J., Fabbri, D., and Bodenheimer, B. (2018). Heatmap Generation for Emergency Medical Procedure Identification.

produced via the same software used to capture medical events in the Center for Experiential Learning and Assessment lab; a trained researcher with a background in pre-hospital medical care observed procedures and documented start and stop times as defined based on work conducted in the controlled environment of the CELA lab.

The main focus for the NFD data has been on using EMG and IMU metrics to classify ESI scores for patients so we might predict ESI scores in the future based on monitoring medics. The ESI triage system is commonly used in civilian emergency department (ED's) and range from most (1) to least (5) severe; patients are assigned an ESI triage score at the time of ED arrival based on the initial assessment of a trained nurse. We were able to capture patients with ESI scores 2-5, but any measure of patient status could have been used.

Using the 22 patient transports for which we have ESI scores, we analyzed features to determine their predictive value to their ESI score. All values are measured from the point in which medical treatment begins within the ambulance (any treatment performed before the patient is brought into the ambulance is ignored) to when medical treatment ends. One feature, the amount of time patient was in the transport, was used to control the other 9 features in the statistical analysis. The features used in this analysis are:

1. Amount of time
2. EMG absolute value mean
3. EMG standard deviation
4. EMG total power
5. EMG DC component (base level of activity)
6. EMG entropy (how erratic muscle movements are)
7. IMU integral modulus acceleration mean (energy expenditure)
8. IMU integral modulus acceleration standard deviation (variance across sections of activity)
9. IMU signal vector magnitude mean (energy expenditure)
10. IMU signal vector magnitude mean (variance across sections of activity)

Results. Using spearman correlation and controlling for the length of the total time medical procedures were being performed during transport, the IMU integral modulus acceleration mean was found to be statistically significant (p-value = 0.000342) and a strong negative correlation (r-score = -0.8, 95% -0.88 to –0.22). Other features showed strong statistical correlation but were not statistically significant in this sample size.

These results indicate the total energy expenditure of a medic correlates with patient acuity measured by ESI triage score at ED arrival, but additional study is required.

Table 3 Sensor data correlation with ESI

Variable	n	r	CI95%	r2	adj_r2	p-value	power
emg_mean	15	-0.28	[-0.69, 0.27]	0.08	-0.074	0.308284	0.178
emg_std	15	-0.2	[-0.64, 0.35]	0.039	-0.122	0.482899	0.109
emg_total_power	15	0.132	[-0.41, 0.6]	0.017	-0.146	0.638744	0.075
emg_dc_power	15	0.132	[-0.41, 0.6]	0.017	-0.146	0.638744	0.075

emg_entropy	15	-0.25	[-0.68, 0.3]	0.064	-0.092	0.361816	0.151
imu_vec_mag_mean	15	-0.8	[-0.93, -0.49]	0.64	0.58	0.000342	0.974
imu_vec_mag_std	15	-0.71	[-0.9, -0.32]	0.51	0.429	0.002774	0.889
imu_int_mod_mean	15	-0.66	[-0.88, -0.22]	0.437	0.343	0.007331	0.806
imu_int_mod_std	15	-0.72	[-0.9, -0.33]	0.52	0.441	0.002399	0.899

Task Decomposition for Clinical Procedure Recognition

The purpose of this effort was to determine if decomposing the clinical procedures into their respective sub-tasks can be used to improve overall clinical procedure recognition via wearable sensors only. The wearable sensors' signal variability within a procedure resulted in confusion across procedures, particularly those of short duration. The decomposition into subtasks focused on using machine learning to identify sub-task elements in order to improve the overall procedure recognition to mitigate the sensor noise. The performed analysis of sub-tasks for recognizing the overall procedure focused on two procedures from each procedure duration: short-, mid- and long-duration. Within these groups the procedure with the best detection rate and the worst detection rates from the prior method were analyzed.

Short-Duration Procedures: CPR – Best, Intravenous medicine administration – Worst

Mid-Duration Procedures: Intubation - Best, King Airway – Worst

Long-duration Procedures: Bagging – Best, IV line – Worst

The Baseline algorithm employed a Procedure RF classifier, which trained an RF classifier to directly predict the procedure from the Myo features. The Baseline algorithm is similar to the context-less classifier that was utilized in prior research.

The new detection methods us a subtask decomposition-based approach, where a subtask RF classifier was trained to predict the class probabilities for 24 subtasks. The twenty-four subtask class probabilities can be interpreted as the confidence with which the Subtask RF classifier classifies each subtask for a given feature window. There is no difference between the three algorithms during the training phase; the difference between the three algorithms exists only during the testing (or inference) phase, based on how the subtasks' class probabilities were utilized to infer the actual procedure.

The Naive Subtask decomposition method assigns equal weights to the subtask probabilities based on the number of subtasks require for a procedure. The Weighted Subtask Decomposition procedure assigns weights to the subtask that are proportional to the time each subtask takes within the procedure. The sequential subtask decomposition procedure looks at the sequence of subtasks within a procedure, thus the temporal relationship between subtasks is leveraged.

Generally, the decompositions did not improve overall prediction accuracy. However, the weighted decomposition method performed significantly better than the naïve subtask decomposition model, by improving overall accuracy and reducing the number of confused sub-tasks. The sub-tasks approach adds new potential confounds to the learning and detection processes, such as subtasks that overlap, subtask that are unique to a particular procedure, and the time required to perform each subtask.

The sequential subtask decomposition method was expected to perform the best, but was actually the worst. One reason for this result was that the method relies on the subtask prediction sequence to classify the procedures. The second reason was that the subtask prediction sequence was processed at a fixed window size at each timestep and did not consider the order of the subtasks.

A limitation of this overall analysis was a limited data set.

Patient Space.

We evaluated the accuracy of the system to correctly classify 24 clinical procedures automatically from video data (Figure 9). Our video data input consisted of a patient and an emergency medical technician (EMT). In this phase of the work, OpenPose was applied to the video data, including 18 different key point positions comprising hands, feet, and head. An example of OpenPose processing of data is shown in Figure 9, in which the EMT places an intrafosseous infusion (IO) line, which is a key pre-hospital procedure. We process the output of the OpenPose data to produce a format we call PatientSpace, which consists of distances between the EMT's two hands and 18 different key locations on the patient.

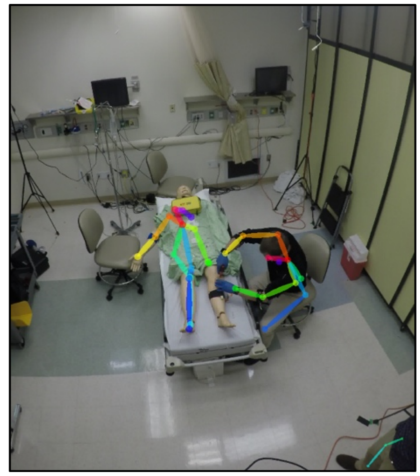


Figure 9. OpenPose output of a frame of data during an instance of the IO line procedure.

We have recorded data for seven subjects who performed 20+ procedures, each multiple times. Using this processed data from seven subjects as feature data, we applied eight machine learning algorithms to the data to classify this data. Data for six subjects was used as training, and the remaining subject was used as testing data. The highest accuracy for any algorithm approached 20%. The neural network algorithm performed best (~18%), followed by gradient boosting. The confusion matrix for the gradient boosting algorithm is shown below. It demonstrates that gradient boosting can reliably identify only about eight of the categories. Future work to increase accuracy of this machine learning classification algorithm approach will include additional subjects performing repetitions of procedures

	ADMI	BAGG	BLOO	CHES	CHES	CLAP	COMI	COMI	CPR	CPR	DRAV	ECG	IM AI	INTU	IO LI	IV LI	IV TO	KING	ORAL	PULS	SPLI	SUTU	SWA	VITA	WRAP
ADMINISTER MEDICATION	0	1509	0	0	0	0	0	206	0	296	0	0	0	0	0	2585	0	0	8	0	703	4416	0	0	0
BAGGING	0	20267	0	0	0	0	0	309	0	1157	0	0	0	0	0	6524	0	0	31	0	216	913	0	0	0
BLOOD-PRESSURE CUFF	0	934	0	0	0	0	1	85	0	208	0	0	0	0	0	1932	0	0	0	0	501	923	0	0	0
CHEST-TUBE	0	1132	0	0	0	0	11	145	0	608	0	0	0	0	0	3779	0	0	0	0	378	831	0	0	0
CHEST-TUBE PREP	0	710	0	0	0	0	2	44	0	445	0	0	0	0	0	2193	0	0	1	0	180	232	0	0	0
CLAP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
COMBAT GAUZE	0	658	0	0	0	0	0	21	0	1167	0	0	0	0	0	1937	0	0	0	0	19	160	0	0	0
COMBAT TOURNIQUET	0	692	0	0	0	0	0	27	0	398	0	0	0	0	0	920	0	0	0	0	66	490	0	0	0
CPR (BREATH)	0	799	0	0	0	0	0	39	0	109	0	0	0	0	0	1400	0	0	0	0	14	247	0	0	0
CPR (COMPRESSION)	0	2194	0	0	0	0	0	98	0	150	0	0	0	0	0	3043	0	0	1	0	34	844	0	0	0
DRAW MEDICATION	0	935	0	0	0	0	0	101	0	391	0	0	0	0	0	1960	0	0	1	0	502	1140	0	0	0
ECG LEADS	0	1033	0	0	0	0	1	112	0	410	0	0	0	0	0	2753	0	0	3	0	153	660	0	0	0
IM ADMINISTRATION	0	895	0	0	0	0	5	40	0	797	0	0	0	0	1	1937	0	0	9	0	222	1055	0	0	0
INTUBATION	0	2264	0	0	0	0	0	41	0	179	0	0	0	0	0	943	0	0	21	0	20	186	0	0	0
IO LINE	0	878	0	0	0	0	0	25	0	63	0	0	0	0	0	1345	0	0	7	0	225	4268	0	0	0
IV LINE	0	1306	0	0	0	0	2	221	0	398	0	0	0	0	0	5949	0	0	13	0	1111	1488	0	0	0
IV TOURNIQUET	0	501	0	0	0	0	0	63	0	151	0	0	0	0	0	2030	0	0	5	0	153	497	0	0	0
KING AIRWAY	0	1054	0	0	0	0	0	3	0	12	0	0	0	0	0	216	0	0	5	0	4	85	0	0	0
ORAL AIRWAY	0	614	0	0	0	0	0	9	0	26	0	0	0	0	0	269	0	0	8	0	14	52	0	0	0
PULSE-OX	0	171	0	0	0	0	5	9	0	18	0	0	0	0	0	167	0	0	0	0	35	61	0	0	0
SPLINTING	0	639	0	0	0	0	0	40	0	762	0	0	0	0	0	2732	0	0	1	0	72	377	0	0	0
SUTURING	0	1759	0	0	0	0	121	170	0	1458	0	0	0	0	0	9748	0	0	1	0	2425	2089	0	0	0
SWAB AREA WITH ALCOHOI	0	1767	0	0	0	0	9	221	0	612	0	0	0	0	0	4667	0	0	7	0	951	2379	0	0	0
VITAL CHECKING	0	900	0	0	0	0	1	66	0	410	0	0	0	0	0	3008	0	0	5	0	169	893	0	0	0
WRAP HEAD WOUND	0	1247	0	0	0	0	0	19	0	230	0	0	0	0	0	698	0	0	16	0	38	312	0	0	0

Figure 10. Confusion Matrix for the Gradient Boosting technique using Patient Space data.

***Machine Learning Event Identification*.**

The automatic event identification system uses wearable sensors, video, and machine-learning to recognize clinical procedures within a controlled environment is presented. The system demonstrated how contextual information and a majority vote method can substantially improve procedure recognition accuracy for each of the procedures listed in Table 4. The wearable sensor data captures arm movements that are representative of a procedure; however, there is a vast array of clinical procedures that need to be detected, which increases the problem's complexity. This complexity is reduced by determining the "active body region" using image processing.

The Myo device is worn on each of the participant's forearm and captures arm movements and muscle contractions via an inertial measurement unit (IMU) and an 8-channel electromyography (EMG) sensor, respectively. Acceleration and orientation data are captured at 50 Hz, while the EMG data is captured at 200 Hz. The Myo automatically calculates the IMU's roll, pitch, and yaw. A five second window, with a one second stride, is applied to each sensor signal. Various window sizes were analyzed, but the five second window produced the best results.

The signal's mean, standard deviation, and max value are calculated for each window and are typical features extracted for activity recognition [3]. Each sensor signal is transformed into the frequency domain using the fast Fourier transform in order to calculate the signal's spectral entropy. Thus, four features are extracted from each sensor signal resulting in fifty-six features per medic hand.

An orthogonal approach to classification using wearable sensor data is to use image processing to track the medic's hands during the clinical procedures. Many procedures are localized to certain areas on a patient's body, making relative hand location an enticing factor. The image-based hand localization system determines the patient's closest limb to the medic's hands for a particular procedure and uses that information for classifier refinement.

During a procedure, assuming the medics hands are proximal to the patient eliminates the need for 2D to 3D image conversion. Thus, the calculated distance between the medic's hand keypoints and each skeleton keypoint on the patient is in pixel space. This measurement's variability and noise is reduced by averaging the limb position over 1 second (24 frames) in order to determine the patient's closest limb to the medic's hands per second. The closest limb is mapped to one of four body regions: head, chest, arm, or leg.

The extracted features from the Myos' IMU and EMG sensors are fed into a random forest classifier, which is a supervisory-based machine-learning algorithm that is an ensemble of individually trained decision tree classifiers. The random forest classifies a signal by taking the class mode of the decision tree ensemble. 100 decision trees with a max-depth of 500 are used for this work, where the parameters were chosen based on classifier performance.

The targeted domain requires knowing if a procedure was performed, not that every single window is correctly classified. Assuming a procedure's start and stop time is known, the procedure can be classified as the majority vote of each classified window within the procedure time frame. For example, if CPR (chest compressions) consists of fifteen windows where ten windows are classified correctly and the other five windows are not, then the procedure can be correctly classified as CPR. Algorithm 1 provides the pseudo code for this classification. The algorithm cycles through each window between the procedure start and stop time, extracting features from the wearable sensor data for each window. DetermineBodyRegion (runs OpenPose on the window's image data and determines the window's active body region, which is used to determine which trained random forest classifier to apply. The extracted features are fed into the classifier to predict a clinical procedure for the window. After each window is processed, the algorithm returns the Majority Vote of the predicted procedures using Max(ProcedureCount()).

Algorithm 1 Clinical Procedure Classification Algorithm

Input: Procedure Start/Stop Time, Wearable Sensor Data, Video Data

Output: ProcedureClassification

```
PredictedProcedureList = []
for each window between Procedure Start and End time do
    Features = ExtractFeatures(window, WearableSensorData)
    ActiveBodyRegion = DetermineBodyRegion(window, Video Data)
    Classifier = DetermineClassifier(ActiveBodyRegion)
    Procedure = Classifier.Predict(Features)
    PredictedProcedureList.append(Procedure)
end for
return Max(ProcedureCount(PredictedProcedureList))
```

The system was validated using leave-one-subject-out cross-validation, where the random forest classifier is trained on two participants' randomly shuffled data and tested on the third participant's data. Five random forest classifiers were trained per cross-validation fold. One classifier was trained using data from every clinical procedure, which represents not knowing the active body region. The other four classifiers correspond to a body region (e.g., head, chest, arm, or leg) and were trained using the respective procedure data. The collected dataset created a class imbalance between procedures, which decreases performance. Thus, the overrepresented procedures are randomly down-sampled during training in order to better balance the class set.

The cross-validation analysis was applied to three conditions: Unknown Body Region, Perfect Body Region, and Estimated Body Region. The unknown body region condition allows for analyzing how the clinical procedure detection system performs without image data (i.e., with only wearable sensor data), while the perfect body region condition assumes that the active body region is always known accurately (i.e., if a procedure corresponds to the head, then the system correctly identifies the head as the active region). The random forest and majority vote methods are analyzed within each body region condition.

Procedure	Body Region Condition					
	Unknown		Perfect		Estimated	
	RF	MV	RF	MV	RF	MV
IO Medication	0.00	0.00	0.05	0.00	0.00	0.00
IV Medication	0.12	0.27	0.37	0.36	0.03	0.00
Bagging	0.43	0.71	0.86	0.86	0.48	0.33
Blood-Pressure Cuff	0.03	0.00	0.39	0.60	0.12	0.50
CPR (Breath)	0.17	0.18	0.30	0.23	0.32	0.66
CPR (Compressions)	0.96	1.00	0.99	1.00	0.21	0.33
Chest-Tube	0.02	0.00	0.42	0.57	0.32	0.66
Combat Gauze	0.37	0.25	0.01	0.00	0.00	0.00
Combat Tourniquet	0.12	0.00	0.52	0.75	0.03	0.00
Draw Medication	0.20	0.20	0.47	0.47	0.32	0.66
ECG Leads	0.12	0.20	0.38	0.40	0.27	0.33
IM Administration	0.03	0.10	0.05	0.10	0.05	0.00
IO Line	0.14	0.29	0.61	0.86	0.15	0.00
IV Line	0.02	0.00	0.22	0.30	0.04	0.00
Intubation	0.27	0.33	0.49	1.0	0.28	0.66
King Airway	0.02	0.00	0.08	0.20	0.02	0.00
Oral Airway	0.09	0.08	0.27	0.33	0.00	0.00
Pulse-Ox Monitor	0.02	0.00	0.48	0.80	0.00	0.00
Splinting	0.13	0.00	0.80	1.00	0.18	0.33
Swab Area with Alcohol	0.00	0.00	0.12	0.13	0.06	0.00
Tie IV Tourniquet	0.03	0.00	0.17	0.11	0.01	0.00
Vital Monitoring	0.71	0.80	0.74	1.00	0.14	0.00
Wrap Head Wound	0.04	0.20	0.39	0.40	0.12	0.33
Average	0.18	0.19	0.40	0.50	0.14	0.21

ECG: Electrocardiogram and **IM**: Intramuscular

Table 4. Classification accuracy (%) by Procedure, Known Body Region Condition, and Classification method: Random Forest (RF) and Majority Vote (MV)

The classification accuracy by procedure and known body region type are presented in Table 4. Overall, CPR (chest compressions) tended to be classified accurately the most, followed by bagging. These accurate classifications were due to the procedures' repetitiveness (i.e., chest compressions or squeezing the bag-valve mask). Vital monitoring was classified accurately as well, due to the procedure requiring minimal arm movements. Short-duration procedures, (i.e., oral airway or swabbing an area with alcohol), were difficult to classify and were often misclassified as a longer-duration procedure. Additional training data will potentially increase classification accuracy for short-duration procedures.

The classification accuracies corresponding to the unknown body region condition serve as a baseline condition, as no contextual data was used. The random forest method and majority vote method achieved an average classification accuracy of 18% and 19%, respectively. The majority vote method increased classification accuracy by at least 10% over the random forest method for five procedures, while two procedure's classification accuracies decreased.

Knowing the active body region with perfect precision increased classification accuracy dramatically for the random forest and majority vote methods, as the methods achieved an average classification accuracy of 40% and 50%, respectively. There was at least a 10% accuracy increase from the unknown body region condition for seventeen procedures using the random forest method and for nineteen procedures with the majority vote method. Both methods experienced a substantial decrease in accuracy for the combat gauze procedure. The majority vote method increased classification accuracy by at least 10% from the random forest method for nine procedures, while no procedure accuracy decreased by more than 10%. These results demonstrate that the majority vote method performs better than the random forest method, when the active body region is correctly identified.

Estimating the active body regions did not change the average classification accuracies dramatically from not knowing the active body region. Six procedures' random forest classification accuracies increased by at least 10%, while five procedures' accuracies decreased by at least 10%. The majority vote method using the estimated body region increased classification accuracy for ten procedures without knowing the body region, while seven procedures' accuracies decreased. Additionally, the majority vote method increased nine procedures' accuracies by at least 10% from the random forest method, while three procedures' accuracies decreased.

Overall, correctly identifying the active body region achieved the highest performance with both classification methods, illustrating the utility of using contextual information in activity recognition. The majority vote method achieved higher average classification accuracies than the random forest method, demonstrating the majority vote method's utility in a real-world complex environment.

Convolutional Neural Networks with Inception Models for Event Identification.

We also attempted to predict the clinical procedure being performed using convolutional neural networks (CNNs) over the video data. Our basis for this task was the work by Karpathy et al. who used CNNs to classify YouTube videos. In this work, the same basic data as for the PatientSpace data was used, except the raw video data was kept, cropped, and resized to 256 x 256. The basic type of data used for this algorithm is shown in Figure 3. This frame data was then sorted according to the paramedic and procedure categories in the 24 categories used for training and analysis.

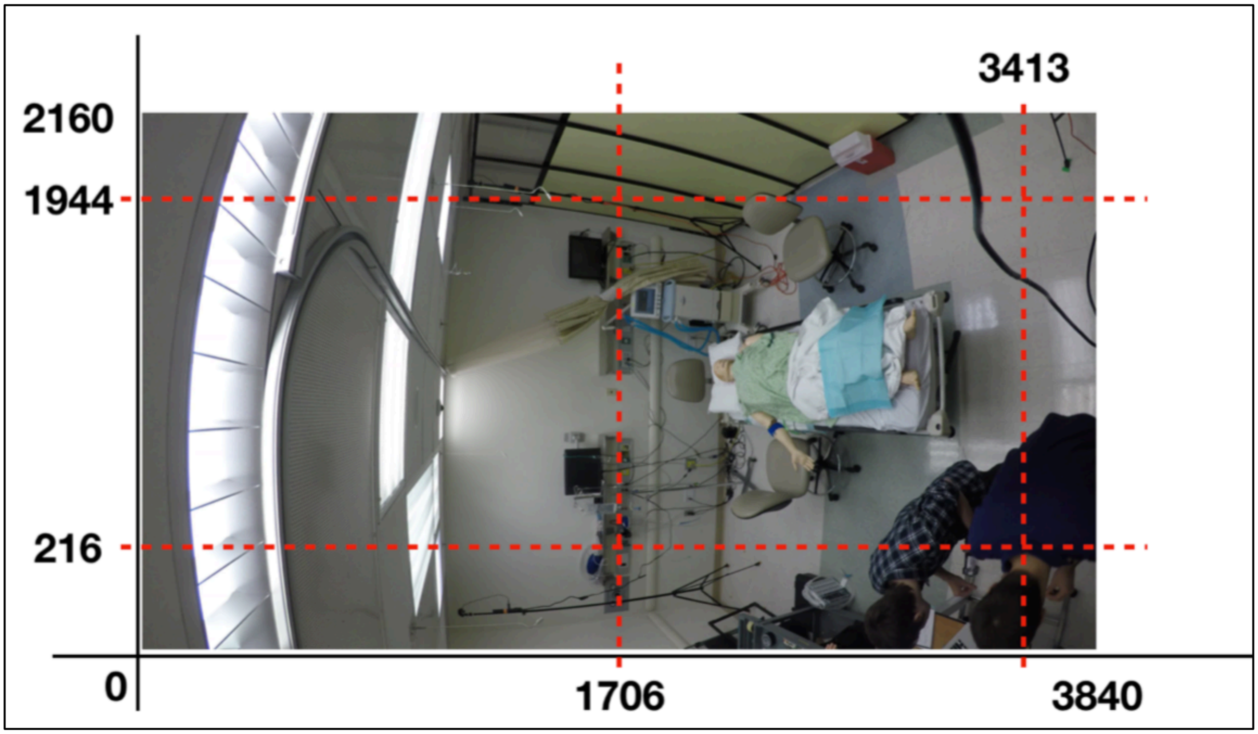


Figure 11. Image cropping area for frame data.

The data for each category is not equal in terms of raw frames, which is how the CNN models will be trained. Categories ranged from 60,000 frames down to 3,000 frames of data. To ensure fairness and balance in training, large categories were subsampled to 10,000 frames and small categories had frames randomly duplicated to increase their frame count to 10,000 frames. Image augmentation was randomly applied to duplicated frames to ameliorate the effects of this cloning. For testing accuracy, the same balancing is applied to insure equal representation across the categories. It should be noted that the training data used in Karpathy et al. was significantly larger by orders of magnitude, and this reinforces the take home message from the section on Patient Space data, that modern deep learning algorithms typically expect or need larger training data.

As in the case of Patient Space data, we have data from seven subjects. For the CNN models, we will use five subjects for training, one for training, and one for validation. Results are currently reported for one-fold validation.

Following the method of Karpathy et al., two CNN architectures are tested, based on the Inception V3 network, as shown in Figure 4. The pre-trained Inception V3 network is used in four forms:

1. Main model. Inception V3 with ImageNet weights for 24 categories of clinical procedures using the single frame CNN approach.

2. Variant 1. Half size Inception V3 with 24 categories using the single frame approach.
3. Variant 2. Main model with 23 categories, combining CPR Breath and CPR compression, as initial results from the main model suggested that it could not distinguish these categories.
4. Late fusion. Inception V3 with ImageNet weights for 24 categories of clinical procedures using the late fusion CNN approach. This approach takes three frames, each 15 frames apart, spanning 1 s of video data in total. Thus, this model summarizes data across a temporal span.

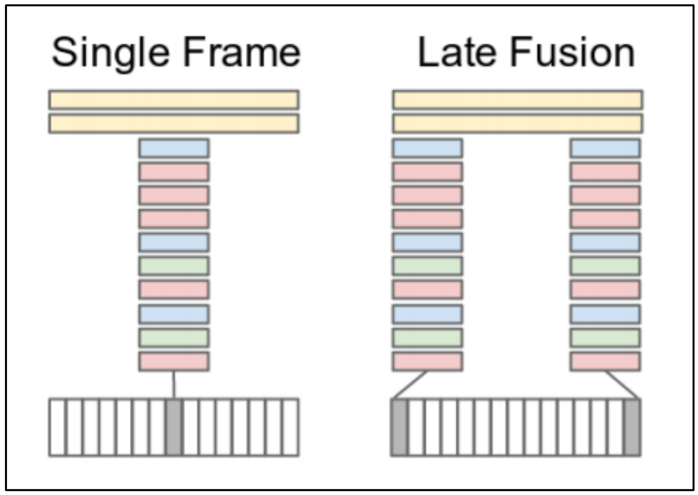


Figure 12. The two explored approaches for fusing temporal frame information into a CNN. Red, green, and blue boxes indicate convolutional, normalization, and pooling layers, respectively. This image taken from Karpathy et al.

The results of this procedure are that all CNN models achieve at least 90% training accuracy. The results for a one-fold validation are that the main model achieves 48% accuracy, variant 1 46% accuracy, variant 2 58% accuracy, and the late fusion model 55% accuracy. Results for five-fold validation are ongoing.

Findings:

1. CNN models with pre-trained weights achieve significantly higher training and validation accuracies on raw data than we have achieved through other means.
2. Data size, particularly sample size across procedure categories, remains a concern.
3. Significant improvement is needed in this area to achieve acceptable and deployable results.

Item Set Pattern Mining for Event Identification.

In contrast to standard machine learning approaches for activity detection, the team also worked to detect activity by looking at temporal motion patterns.

Each participant wore two Myo devices in the left and right hand. The channel number is from 1 to 8 in the left hand, and 9 to 17 in the right hand. We define an item as the EMG activity signal in a specific EMG channel at a specific time point. We define an itemset as the EMG activity signal from all EMG channels at the same time point. We represent a clinical procedure with sixteen sequences of itemsets from all myo EMG channels. As shown in Figure 6, all EMG signals in time stamp 0 form an itemset: {EMG1_1=1, EMG1_2=-1 … EMG2_8=1}.

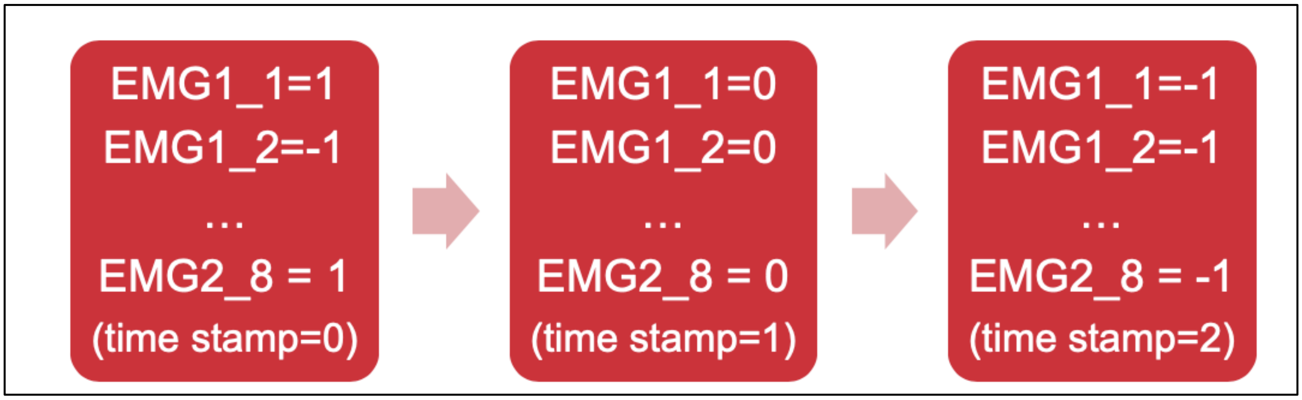


Figure 13. An example sequence of EMG signal itemsets. (EMG1_* represent the signals from the left hand, and EMG2_* represent the signals from the right hand.)

We define a frequent itemset as an EMG signal subset that occurs in all the participants (i.e., we set the minimum support ratio to 100%) of the same procedure. As shown in Figure 2, the frequent length-1 itemsets are (EMG1=1), (EMG2=1), and the frequent length-2 itemsets (EMG1=1, EMG2=1), which occur in all participants. We define a frequent itemset with length-K as a frequent K-itemset.

participant 1 = {(EMG1=1, EMG2=1), (EMG1=1, EMG2=-1), (EMG1=1, EMG3=1)}
participant 2 = {(EMG1=1, EMG2= 0), (EMG1=1, EMG2=1), (EMG1=1, EMG3=0)}
participant 3 = {(EMG1=1, EMG2=1), (EMG1=1, EMG2=0), (EMG1=1, EMG3=-1)}

Figure 14. Example sequences of EMG signals and associated itemsets.

We define a critical K-itemset Ck as the one that embeds information such as:

1. A procedure that contains Ck must be a specific procedure or a list of specific procedures.

2. A procedure that contains Ck must not be a specific procedure or a list of specific procedures. Table 5 shows the critical itemsets and the procedures they exclude when existing. The critical itemset “5_1” reveals that if we monitor positive signal values in left-hand EMG channel 6, then the procedure must not be “Administer IO Medication.” Similarly, the critical itemset “5_1->14_1” reveals that if we monitor negative signal values in the left-hand EMG channel 6 and the right-hand EMG channel 15, then the procedure must not be “Administer IO Medication” and “Administer IV Medication.”

Critical Itemset	Excluded Procedures When the itemset Exist
5_1	Administer IO Medication
2_-1->8_1	Bagging
0_-1->7_-1	Administer IM Medication,

	Administer IO Medication, Administer Medication, Bagging, CPR (Breath), Chest-Tube Preparation, Intubation, Suturing, Vital Checking
5_1->14_1	Administer IO Medication, Administer IV Medication

Table 5. Example Critical Frequent 1-Itemsets and 2-itemsets.

Evaluation of Frequent Itemsets. We evaluate the predictive power of the frequent itemsets in classifying all clinical procedures. We evaluate different subsets of frequent K-itemsets, which K is from 1 to 9. Table 6 shows the average accuracy with different K values. The result shows that the combined frequent 4 & 5-itemsets provide the highest average accuracy. Figure 14 shows the result's confusion matrix when using the combined frequent 4 & 5-itemsets.

Frequent Itemset	Average Accuracy
1-itemsets	0.13
2-itemsets	0.22
3-itemsets	0.26
4-itemsets	0.35
5-itemsets	0.35
4-itemsets & 5-itemsets	**0.40**
6-itemset	0.21
7-itemset	0.13
8-itemset	0.04
9-itemset	0.04

Table 6. Average Prediction Accuracy using different Frequent Itemsets.

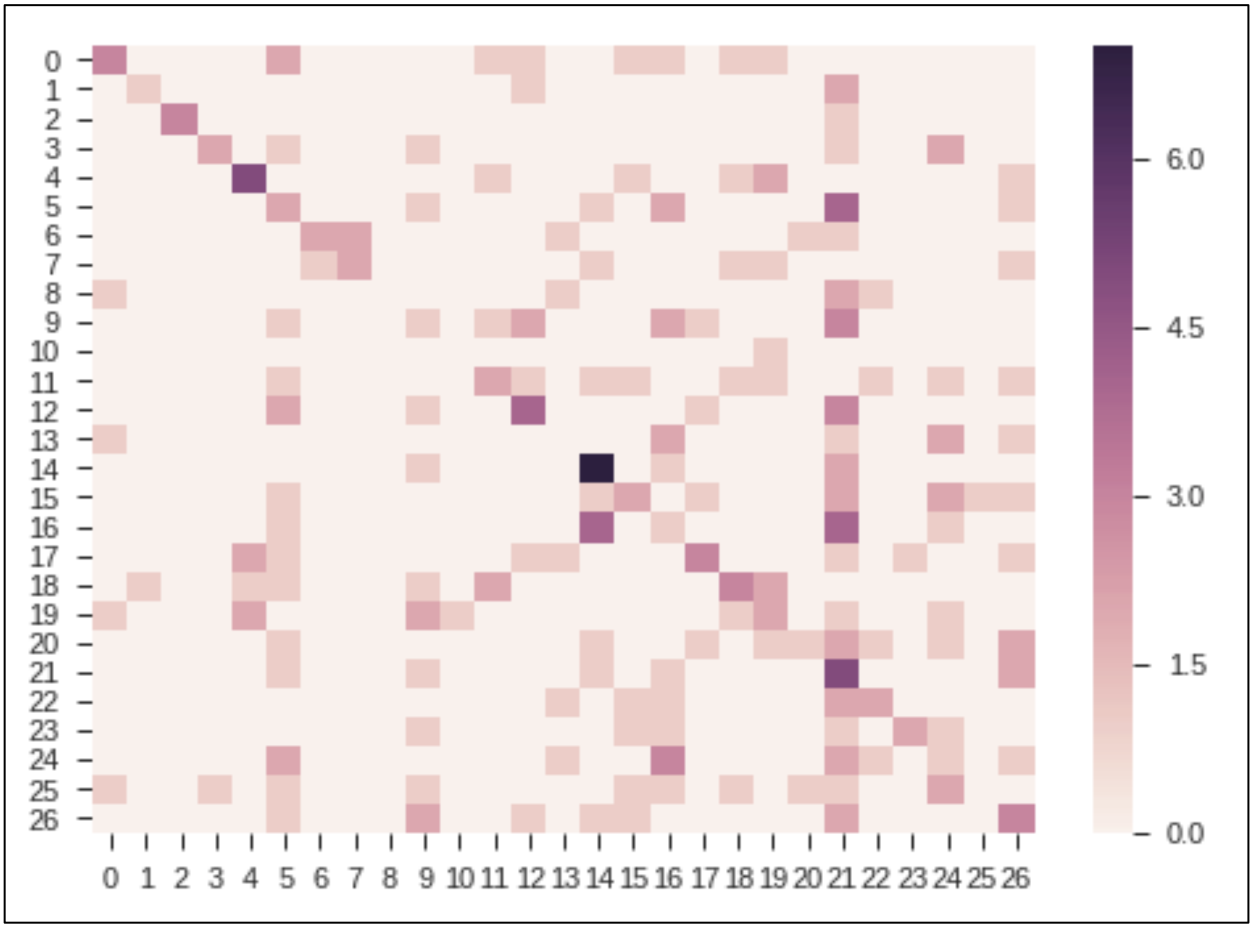


Figure 14. Confusion Matrix when using the 4-itemsets & 5-itemsets (40% accuracy).

Conclusion: Activity patterns demonstrate their ability to predict clinical procedures without leveraging the body area information or binning the procedures. More analysis and development are needed to identify more powerful activity patterns.

Emergency Clinical Procedure Detection via Wearable Sensors

Communicating a patient's state accurately during transfer from emergency medical technicians to hospital personnel is crucial for optimal care. Prior work demonstrated automated algorithms that combined wearable sensors with video data from cameras to detect clinical procedures and improve this information transfer. However, incorporating video requires task- or environment-specific installation mechanisms, raises privacy concerns, and is susceptible to occlusion and image noise. The presented approach detects clinical procedures using wearable sensors (i.e., inertial and electrophysiological) only and the procedures' subtasks to mitigate the sensors' signal variability to provide clinical procedure detection.

A paper was submitted for review in March 2021 to the Human Factors and Ergonomics Society Annual.

Field Testing Among Nashville Fire Department Paramedics.

Data collection was paused due to COVID-19 and we were not able to acquire sufficient samples.

Study participant NFD paramedics continue to show a high level of engagement and excitement about the project. They have verbalized that they enjoy participating in VUMC research and would like to continue to be a part of the study.

Leadership at NFD has been accommodating before COVID-19 by allowing us to self-schedule observations with dates/times that are most convenient for the observer and were supportive of press releases related to this collaboration (Appendix N).

4. IMPACT:

Development of the principal discipline(s) of the project. The goal of the project is to leverage off-the-shelf sensors to automatically generate electronic health record documentation that can enable trauma team preparedness. In the broader field of clinical documentation, individual sensors have been previously used in isolation to measure vital signs and detect simple activities in a variety of healthcare settings. The project, through its aim of using sensors to create a unified data feed of clinical care during patient transport, identify interventions and produce a triage score, will extend current approaches to clinical documentation to generate, automatically, new information. The techniques that the project is employing, which are well suited to environments where hands-free data entry is essential, has the potential to drastically improve the performance of medical teams by creating more robust communication pathways. Additionally, because the approach is not voice dependent, it has the potential to create new opportunities for hands free data collection in environments where ambient noise prevents the use of voice technologies.

Other disciplines. The project's approach is likely to impact approaches to automatic task data collection in fields outside of healthcare. While our focus is to use multi-sensor technologies to detect clinical tasks in the operational environment, once refined the same approaches could be used to detect other tasks in a variety of settings. For example, this approach could be modified to create the ability to detect whether an individual has performed an equipment check out procedure, completed a set of activities in a particular order, or any other set of tasks where the surveillance of a particular activity can provide useful information.

Technology transfer. While project's technology will enable deployed medical personnel to coordinate care more effectively among combat casualties, it also will have the ability to transfer within the civilian sector during emergency medical trauma treatment. The intent of the project is to widely distribute its methodology as we foresee these techniques being used in both the military deployed environment and the civilian pre-hospital sector. Development of a handsfree system to automatically generate and transmit a clinical care record will bridge the gap between current communication and documentation practices so that information can flow seamlessly and in real-time across settings of care in almost any environment – military or civilian.

In 2021, a patent application was submitted to the US Patent Office.

Society beyond science and technology. The project's approach is likely to impact society beyond the bounds of science through the ability of these techniques to help improve social conditions. We expect the methodologies developed will enable rapid and measurable contributions to public health and economic output through the ability of a variety of processes to be enhanced, improved, and adjusted to run more efficiently. In healthcare, we expect patients will benefit. In manufacturing, we expect supply chains could be improved. In the military, we expect a variety of operational platforms could be enhanced.

5. CHANGES/PROBLEMS:

On December 24, 2019, the Department of Defense provided written approval the request for a no-cost extension (NCE) to December 29, 2020. It states that "the purpose of this modification is “to execute a no cost extension and update the SOW to reflect the extended period of performance. The total contract value is unchanged. The period of performance is changed.”

Our main focus for 2020 was finishing data analytics tasks, and also collecting data with the Nashville Fire Department. Unfortunately, due to **COVID-19**, data collection was halted and has not resumed, thus limiting data available for analysis.

6. PRODUCTS:

A. <u>Papers:</u>

1. Novak, L. L., Simpson, C. L., Coco, J. R., McNaughton, C. D., Ehrenfeld, J. M., & Fabbri, D. (2020). Understanding the Information Needs and Context of Trauma Handoffs to Design Automated Sensing Clinical Documentation Technologies: Qualitative Mixed-Method Study of Military and Civilian Cases. Journal of Medical Internet Research.

B. <u>Conference Papers:</u>

1. Li, L., Paris, R., Pinson, C., Wang, Y., Coco, J.., Heard, J., Adams, J., Fabbri, D., Bodenheimer, B. (2020). Emergency Clinical Procedure Detection with Deep Learning. IEEE Engineering in Medicine & Biology Society.

2. Heard, J., Paris R., Scully, D., McNaughton, C., Ehrenfeld, J., Coco, J., … Adams, J. (2019). Automatic Clinical Procedure Detection for Emergency Services. IEEE Engineering in Medicine and Biology Society.

3. Paris, R., Sullivan, P, Heard, J, …, Ehrenfeld, J., Bodenheimer, B. (2019) Heatmap Generation for Emergency Medical Procedure Identification. SPIE Medical Imaging.

4. Bloos, S. M. (2019). Feasibility Assessment of a Pre-Hospital Automated Sensing Clinical Documentation System. American Medical Informatics Association (AMIA).

   ** The oral presentation received positive press, to include:*

   A. "Vanderbilt Researchers Test mHealth Platform to Improve ED Hand-Offs" By Erick Wicklund, MHealthInteligence, 27 NOV 19 https://mhealthintelligence.com/news/vanderbilt-researchers-test-mhealth-platform-to-improve-ed-hand-offs

   B. "Documentation system seeks to improve paramedic-ED patient handoffs" By Greg Slabodkin, HealthData Management, 2 DEC 19 https://www.healthdatamanagement.com/news/documentation-system-seeks-to-improve-paramedic-ed-patient-handoffs

   C. "Nashville FD, medical center testing automated documentation system for ER handoffs"
   By News Staff, EMS1.com, 2 DEC 19
   https://www.ems1.com/technology/articles/nashville-fd-medical-center-testing-automated-documentation-system-for-er-handoffs-5tUXEnrafJAqGv1k/

C. *Patents:*

1. Automatic Sensing for Clinical Decision Support. Application Number: 17203204. Filed: 16-March-2021. First Named Inventor: Daniel Fabbri.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS:

Individuals who have worked on the project:

Name:	Daniel Fabbri, PhD
Project Role:	Principal Investigator
Researcher Identifier:	NA
Nearest person month worked:	2.40 calendar months (20% effort/12 months)
Contribution to Project:	Leads the overall research and the project's computation efforts. Chairs project planning meetings consisting of all key and other participating personnel.

Name:	Julie Adams, PhD
Project Role:	Site PI/Co-Investigator, Oregon State University
Researcher Identifier:	NA
Nearest person month worked:	1.63 academic months (14% effort/12 months)
Contribution to Project:	Overseas project, mentored students, provided direction and feedback for research tasks, and prepared deliverables.

Name:	Robert Bodenheimer, PhD
Project Role:	Site PI/Co-Investigator, Vanderbilt University
Researcher Identifier:	NA
Nearest person month worked:	1.20 academic months (10% effort/12 months)
Contribution to Project:	Develops, maintains, and upgrades tracking algorithms and codebase, and incorporated deployable and testable platforms.

Name:	Laurie Novak, PhD, MHSA
Project Role:	Co-Investigator
Researcher Identifier:	NA
Nearest person month worked:	1.20 calendar months (10% effort/12 months) Contribution to Project: Leads the development of the information needs taxonomy.

Name:	Candace D McNaughton, MPH MD PhD
Project Role:	Co-Investigator, Emergency Medicine
Researcher Identifier:	NA
Nearest person month worked:	0.60 calendar months (20% effort/3 months)
Contribution to Project:	Advises project on emergency medicine and
taxonomy	

Name:	Joseph Coco
Project Role:	Health Systems Analyst/Programmer
Researcher Identifier:	NA
Nearest person month worked:	3.84 calendar months (32% effort/12 months)
Contribution to Project:	Assists with prototype software development.

Name:	Christopher Simpson
Project Role:	Research Assistant
Researcher Identifier:	NA
Nearest person month worked:	1.56 calendar months (13% effort/12 months)
Contribution to Project:	Conducts interviews, conducted qualitative data management and analysis, and provided research support.

Other organizations that were involved as partners:

Organization Name:	Oregon State University
Location of Organization:	Corvallis, Oregon
Partner's Contribution to Project:	Collaboration with Electrical Engineering and Computer Science Department

Organization Name:	Vanderbilt University
Location of Organization:	Nashville, Tennessee
Partner's Contribution to Project:	Human simulation utilizing state-of-the-art mannequins.
Facilities:	Center for Experiential Learning and Assessment (CELA)

Organization Name:	U.S. Army Rascon School of Combat Medicine,
Location of Organization:	Fort Campbell, Kentucky
Partner's Contribution to Project:	Hosted visit to better understand Army's Tactical Combat Casualty Care (TCCC)
Facilities:	Medical Simulation Training Center (MSTC)

8. APPENDICES:

A. Project Statement of Work (Updated)

B. Focus Group Interview Guide

C. Key Domain Findings from Focus Groups:

D. Findings from Video Observations

E. Clinical Emergency Procedure List

F. EMG and Accelerometer Graphs Examples

G. CELA Lab Setup, Process, and Data Collection

H. Hierarchal Task Analysis

I. Paper: Heatmap Generation for Emergency Medical Procedure Identification

J. Paper: Feasibility Assessment of a Pre-Hospital Automated Sensing Clinical Documentation System

K. Paper: Understanding the Information Needs and Context of Trauma Handoffs to Design Automated Sensing Clinical Documentation Technologies: Qualitative Mixed-Method

L. Paper: Automatic Clinical Procedure Detection for Emergency Services

M. Paper: Emergency Clinical Procedure Detection with Deep Learning

N. Nashville Fire Press Release

APPENDIX A

Project Statement of Work (Updated)

STATEMENT OF WORK – 12/19/2019
PROPOSED START DATE Sep 30, 2017

Site 1: Vanderbilt University Medical Center
3319 West End Ave, Suite 970
Nashville, TN 37203
Initiating PI: Dr. Daniel Fabbri

Site 2: Vanderbilt University
110 21st Avenue, Suite 800
Nashville, TN 37203
Partnering PI: Dr. Robert Bodenheimer

Site 3: Oregon State University
308 Kerr Administration Building
Corvallis, OR 97331
Partnering PI: Dr. Julie Adams

Specific Aim 1: Design and implement a clinical activity detection prototype system using accelerometers and image data	**Timeline**	**Site 1** (Initiating PI)	**Site 2** (Partner PI)	**Site 3** (Partner PI)
Major Task 1: Develop clinical activity detection algorithms utilizing accelerometer data	Months			
Subtask 1: Design and evaluate methods to extract accelerometer data from devices in real-time or near real-time. • Dr. Fabbri's team oversees the design • Dr. Adams' team implements design and performs data validation	1-4	Dr. Fabbri		Dr. Adams
Subtask 2: Design and implement basic activity detection algorithms using captured accelerometer data • Dr. Fabbri's team develops a taxonomy of clinical tasks to identify • Dr. Fabbri's team collects sample accelerometer data in the simulation lab • Dr. Adams' team implements and evaluates activity detection algorithms on captured simulation data, and test data sets.	4-9	Dr. Fabbri		Dr. Adams
Subtask 3: Aggregate accelerometer data into a centralized physician dashboards • Dr. Fabbri's team develops the data architecture and storage system for the data • Dr. Adams' team implements a system to transmit	9-14	Dr. Fabbri		Dr. Adams

data from accelerometers to the dashboard				
Major Task 2: Develop clinical activity detection algorithms that utilizes image (video) data				
Subtask 1: Image capture and body position detection • Dr. Fabbri's team oversees the design • Dr. Bodenheimer's team designs and implements the image capture system	1-4	Dr. Fabbri	Dr. Bodenheimer	
Subtask 2: Design and implement a basic clinical activity detection system using image (video) data • Dr. Fabbri's team develops a taxonomy of clinical tasks to identify • Dr. Fabbri's team collects sample image data in the simulation lab • Dr. Bodenheimer's team designs and evaluates activity detection algorithms using the image and video data.	4-9	Dr. Fabbri	Dr. Bodenheimer	
Subtask 3: Aggregate image data into a centralized physician dashboards • Dr. Fabbri's team develops the data architecture and storage system for the data • Dr. Bodenheimer's team implements a system to transmit data from cameras to the dashboard	9-14	Dr. Fabbri	Dr. Bodenheimer	
Major Task 3: Develop clinical activity detection algorithms that combines accelerometer and image (video) data				
Subtask 1: Develop models to correlate accelerometer and image data for activity detection.	9-12	Dr. Fabbri	Dr. Bodenheimer	Dr. Adams
Subtask 2: Evaluate combined activity detection algorithms on simulation lab data. • Dr. Fabbri's team will lead the evaluation, but provide feedback to the other teams • Dr. Bodenheimer's team will evaluate the impact of image data for combined activity detection • Dr. Adams' team will evaluate the impact of accelerometer data for combined activity detection	12-36	Dr. Fabbri	Dr. Bodenheimer	Dr. Adams
Subtask 3: Integrate the combined activity detection system into physician dashboard	15-20	Dr. Fabbri		

Major Task 4: Design and implement high-level clinical activity features				
Subtask 1: Design and implement an injury heatmap visualization • Dr. Bodenheimer's team will use image data to determine hand position over the body • Dr. Adams' team will team will use accelerometer data to determine duration of hand positions and movements. • Dr. Fabbri's team will aggregate the data into a website visualization	8-14	Dr. Fabbri	Dr. Bodenheimer	Dr. Adams
Subtask 2: Design and implement a risk score derived from accelerometer and image data • Dr. Fabbri's team will utilize the heatmap data to produce a risk score	14-18	Dr. Fabbri		
Subtask 3: Design and develop 'quick' data entry systems using accelerometer devices (e.g., watches) • Dr. Fabbri's team will oversee the design. • Dr. Adams will implement a 'quick' data entry prototype.	18-24	Dr. Fabbri		Dr. Adams
Major Task 5: Focus group and field data collection of developed prototype systems				
Subtask 1: Focus group of medics, surgeons, emergency room physicians, and military staff.	2-4 & 18-20	Dr. Fabbri (N=30)		
Subtask 2: Field data collection with Nashville Fire Department	13-22	Dr. Fabbri (N=40)		
Milestone #1: Prepare publication on problem overview, and detection algorithms	15-36	Dr. Fabbri	Dr. Bodenheimer	Dr. Adams
Milestone #2: Package prototype software for sharing and distribution	20-36	Dr. Fabbri	Dr. Bodenheimer	Dr. Adams

APPENDIX B

Focus Group Interview Guide:

When thinking about handoffs between pre-hospital and hospital providers who care for trauma patients:

1. What is your impression of handoffs for trauma patients?

2. What information is normally shared during hand-offs? (i.e., essential care information)

    a. What information is most useful to determine next steps in care management? i. Vitals ii. Injury severity iii. Injury location iv. Patient overall state v. Medical history vi. Changes in vitals over time vii. Procedures performed during transport

    b. Why/how is this information shared?

    c. What information is not useful to determine care management?

    d. How might receiving this information before patient arrival assist care management?

3. What issues, if any, have you observed about handoffs for trauma patients? (i.e., what gaps in hand-off reports commonly occur?)

    a. Example of handoff(s) that went well?

    b. Example of handoff that did not go well?

    c. Examples of miscommunications/misunderstandings?

    d. Examples where pre-hospital transport communication was murky, or not well communicated?

    e. Examples of important information communicated too late?

    f. Examples of information that could have been communicated clearer/earlier that would have improved the care/patient outcome?

4. Why/how was this information not shared? (e.g. Due to time constraints, noise)

    a. Type of transfer and mode of arrival (e.g., from scene by ground, air; transfer from other facility by air or ground; if ground, NFD vs. non-NFD)

    b. Time of day?

    c. ESI score?

    d. En Route Report?

APPENDIX C

Key Domain Findings from Focus Groups:

- Vital Signs and Demographics.
  - Vital signs drive a lot of ER's levels, and they are important predictors.
  - The two vital sign values most important for the EMTs and receiving hospital are: 1) the lowest worst number they got, and 2) what it is right now.
  - Vital signs that EMTs capture and/or provide include: Hypertensive, Hypoxic, Heart rate, blood pressure, oxygen saturation, pain, age, meds history, allergies, entitle CO2, O2 saturations.
  - Demographics are important, and include: time of injury, transportation length/time since "uncertainty grows as the time shortens", and age is one of the most important variables.
- Medication Administration
  - The administration (timing and amount) of medicine by EMTs is important because it will change the care provided.
  - It is estimated that a large percentage of field personnel might be guessing after the fact on the actual administration time. On timing, the EMTs are approximate, i.e. within 5 minutes.
  - Medicine provided is in colored boxes. It would be convenient for EMTs for it to be automatically captured upon administration; maybe with a QR code that passes through a camera.
  - Some medicine administered during transport include: Sedative, analgesics, paralytics, Ketamine, Tolinase, Succinylcholine, Rocuronium, Zofran, Epi-drips.
- Documentation
  - EMTs often do not formally document procedures (ie vitals, things they do). They use alternative methods instead using shorthand for their own reference. This includes writing on: Sheets, gloves, paper, tape.
  - EMT documentation include: patient contact time, patient weight, medications given and timing, initial vitals, any procedures that were done, record the number of attempts, the time of those events.
  - EMTs do have a more formal record in the "monitor" where they can document and pull later. This monitor is cumbersome especially when having to scroll through it. It does not transmit in real time to the receiving hospital.
  - Most often, EMTs conduct their formal documentation post-trip arrival. They do like this charting, which can take 30 minutes.
  - There are multiple people in transport, i.e. two people in ground ambulance, and 3-4 four in air transport. This can create challenges in documenting all the procedures.
  - An iPad, or something automated, would provide real value. It will not be that EMTs would go over and show it to the physician, but it would be a quick reference where they could go.
- Transport

- Doctors care about how much and what the medic is doing during transport. EMTs perform routine procedures to include: transfusions, airway, chest tubes, breathing, circulation, full assessment, including vital signs, breathing sounds.
- The receiving hospital may allocate resources in advance based on procedures conducted during transport.
- Sequence matters since it is like cause and effect. They want to know what they did and when it happened as opposed to just a set of collapsed procedures that you do not provide why it happened.
- EMTs want to get the patients in the vehicle and do whatever you can en route. They need a reason to delay transport to give something in the field. They want to get them in the aircraft as quickly as possible and do what they can in flight while in motion towards what they really need, which is a surgeon.

- Communication
  - There are multiple periods of verbal communication between the EMTs and receiving hospital before the patients' arrival, this includes communicating to hospital 10 minutes out.
  - EMTs believe that the doctors do not want too many details when you are at bedside. EMTs just normally tell them what happened. They care less about the individual times and more about if you followed the ACLS algorithm.
  - EMTs need to optimize information to minimize your communication for the doctor so they can take what is important faster. It would be helpful if there was an ability to gather all this information, and let the doctor decide what information he might need most since they want certain things differently.
  - EMTs level their trauma patients en-route which doctors can make inferences and coordinate resources against (i.e. trauma bay, etc.).
  - Communication is difficult whenever there are multiple arrivals back-to-back. Communication must go through Wi-Fi, they have to take the report, it needs to go to the nursing staff, and resources need to be allocated.
  - Due to environmental distractions, doctors might miss a lot of the details of the mechanism, the time of injury, that type kind of thing. Then I would have to go back and try to piece together those details.
  - A few doctors assess that 20% or 30% of the time it might be the wrong information due to perception from what the EMTs are seeing and the dynamic situation. Conversely, what seems like a miscommunication is actually just a change in the patient's condition too.
  - Sometimes the doctors find out they missed or did not get the worst vital so after the patient arrives, they need to upgrade the trauma and re-allocate resources.
  - The biggest lack of communication might be for those patients who are transferred from another facility when the crew just does not have a good report from the outside facility.
  - There is some disagreement that patients arrive frequently with procedures done (i.e. intubation) that is not communicated before they arrive.
  - EMTs admit that they due forget some things such as the number of times something has been done or the severity, or it might not be as clear from memory. It also could be due to the division of labor during transport.

- Having some sort of electronic tracking would be beneficial. Reducing error and reducing communication will improve patient care. Sometimes what the doctor hears and writes down might be different than what the EMT had indicated.
- Doctors encouraged us to build a system where it would pick up on particularly bad situations and create an automatic alert for the hospital.

• Incorporating (new) Technology
  - Doctors trust sensors after they have been analyzed by a person. Regarding equipment, EMTs said that they would “trust but verify”.
  - EMTs wear watches and carry and use cell phones during operations.
  - EMTs are hesitant and nervous to be videoed or recorded.
  - Some EMTs think that it would be nice to have an Apple Watch with a button that you could scroll and hit, especially on busy flights, to know what and when procedures were completed. Other EMTs think that scrolling might get frustrating.
  - Having something take notes without a pen either verbally or with hand gestures would be useful especially due to time compression (ie a lot happening during short flights).
  - It would be great to show the information graphically, and sync with the monitor to automatically put those times in there, and then used when conducting formal documentation.
  - EMTs do not want to have to readjust something under flight suit, which are long sleeves. Pressure of the device might be an issue.

APPENDIX D

Findings from Video Observations

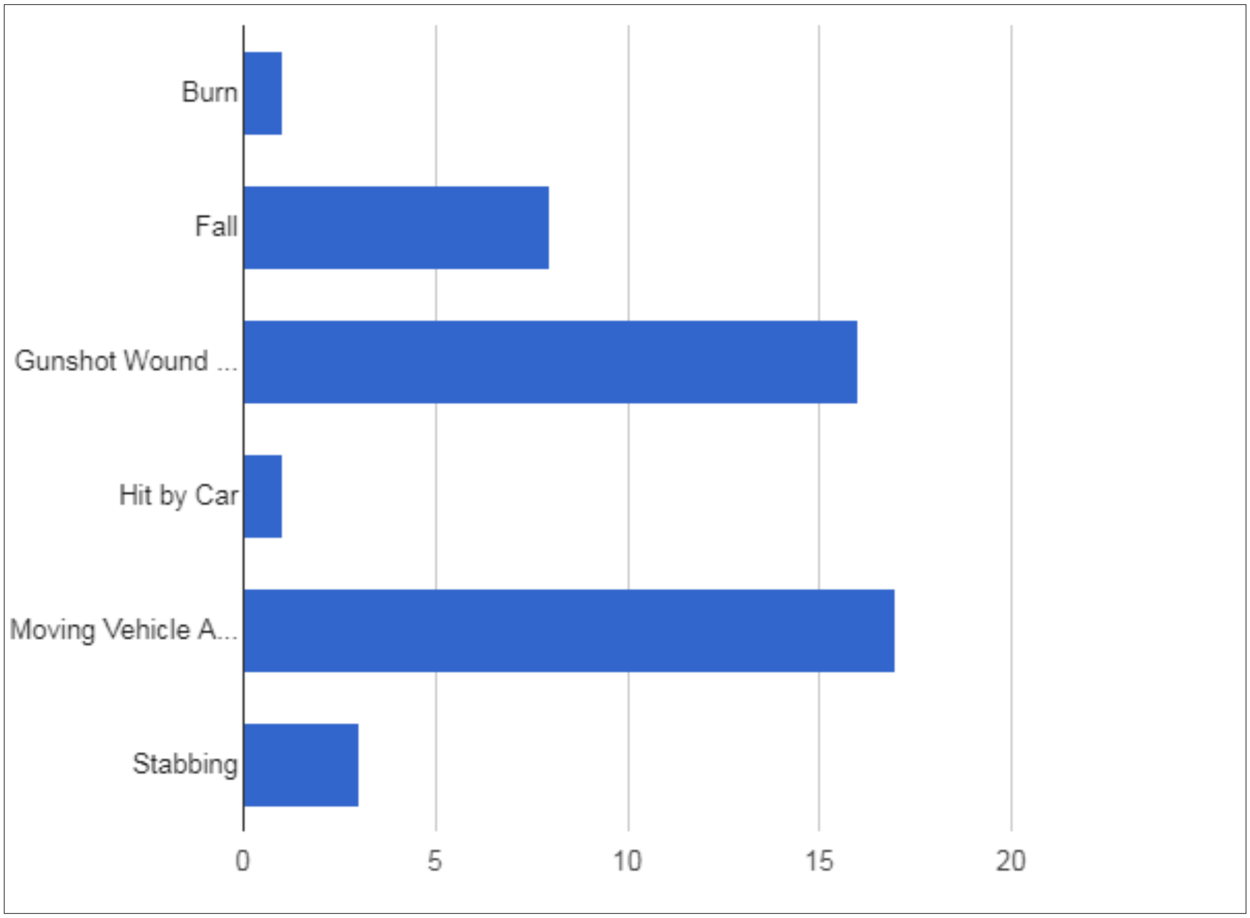


Figure 9. MOI occurrences from videos observed. 4 of the videos contained injuries that could best be described as “other”, including an explosion, and assault, one that was unclear. One of the four videos that were described as “other” injuries were recategorized as a GSW. It was described as a ballistic injury in the video comments.

Counts/frequency: Burn (1, 2.1%), Fall (8, 17.0%), Gunshot Wound (GSW) (17, 36.2%), Hit by Car (1, 2.1%), Moving Vehicle Accident (MVA) (17, 36.2%), Stabbing (3, 6.4%)

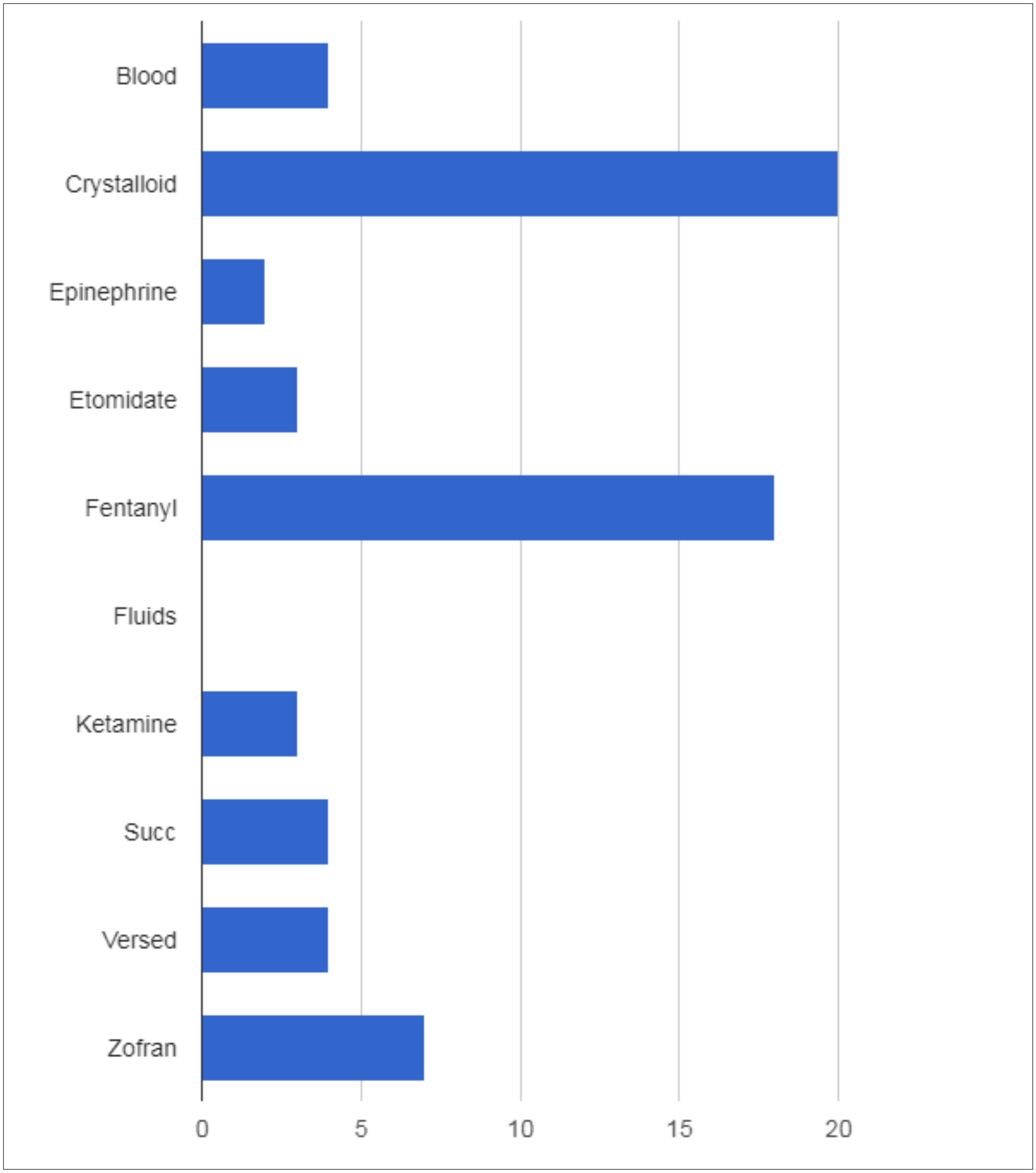


Figure 10. This Figure illustrates the information transferred to the hospital trauma team regarding medications administered during transport. There were 18 out of the 50 videos which did not provide this information.

Counts/frequency: Blood (4, 12.5%), Crystalloid (20, 62.5%), Epinephrine (2, 6.3%), Etomidate (3, 9.4%), Fentanyl(18, 56.3%), Fluids (0, 0.0%), Ketamine (3, 9.4%), Succ (4, 12.5%), Versed (4, 12.5%), Zofran (7, 21.9%)

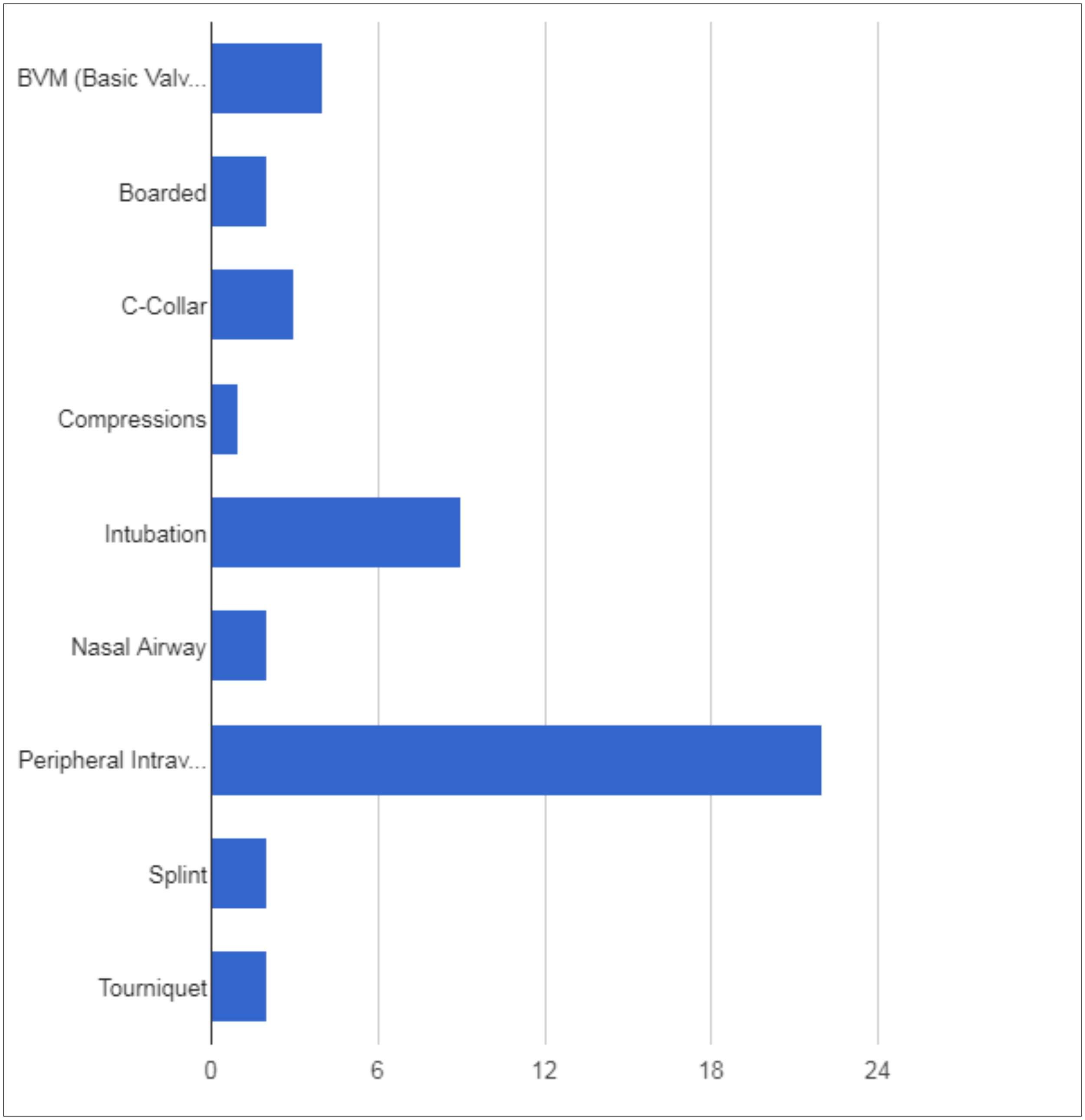


Figure 11. This Figure illustrates the information transferred to the hospital trauma team regarding information about procedures performed during transport. 47 of the 50 cases provided this information.

Counts/frequency: BVM (Basic Valve Mask) (4, 12.1%), Boarded (2, 6.1%), C-Collar (3, 9.1%), Compressions (1, 3.0%), Intubation (9, 27.3%), Nasal Airway (2, 6.1%), Peripheral Intravenous (PIV) (22, 66.7%), Splint (2, 6.1%), Tourniquet (2, 6.1%)

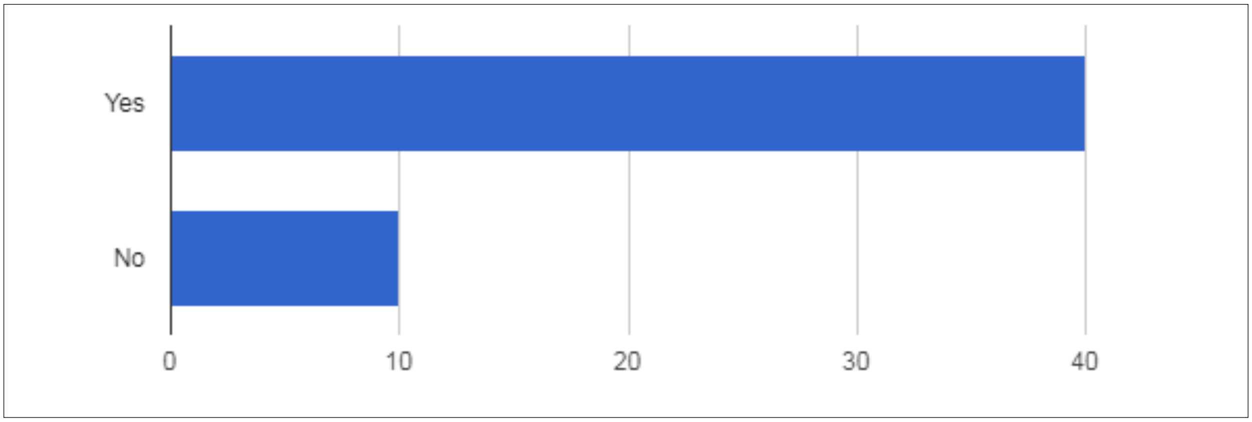


Figure 12. This Figure illustrates that out of the 50 videos, 40 of them contained clarifying questions from the receiving hospital trauma team.

Counts/frequency: Yes (40, 80.0%), No (10, 20.0%)

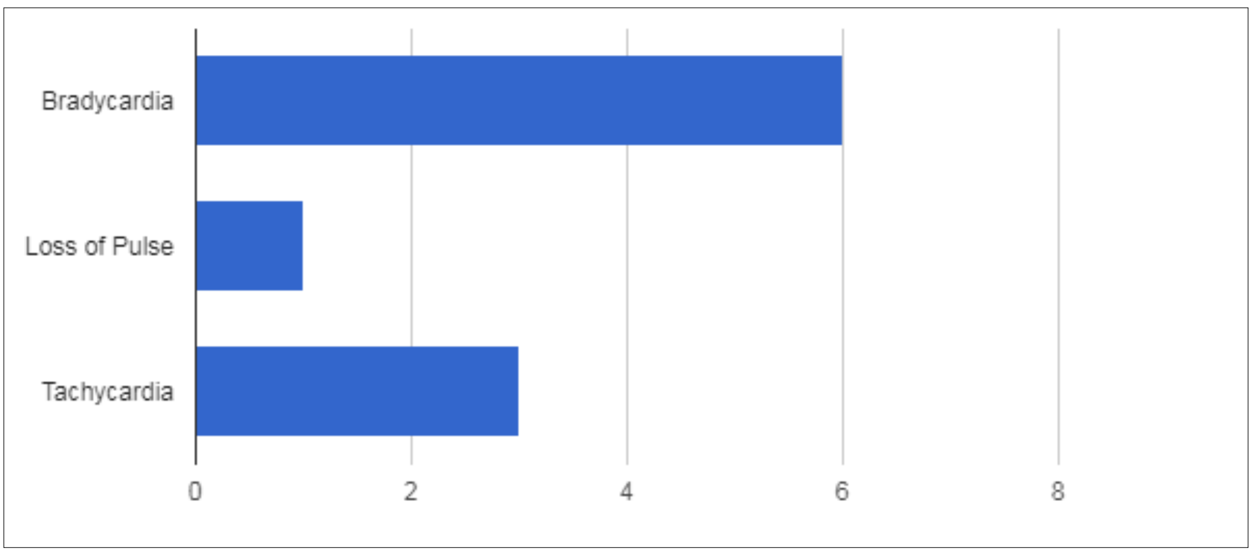


Figure 13. This Figure illustrates the fact that out of the 50 videos observed, information about heart rate was only relayed to the hospital trauma team in 10 of the videos.

Counts/frequency: Bradycardia (6, 60.0%), Loss of Pulse (1, 10.0%), Tachycardia (3, 30.0%)

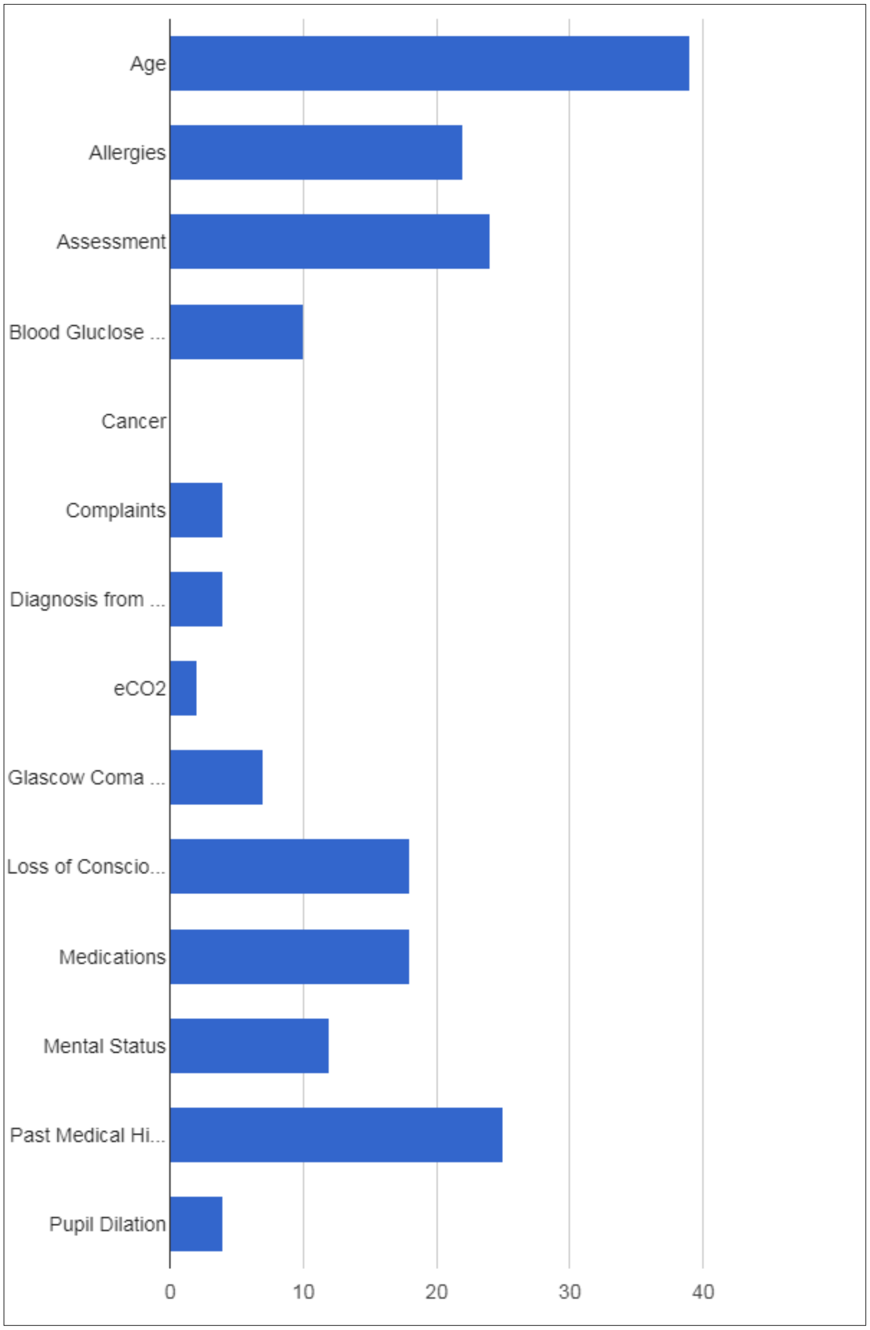


Figure 14. This Figure illustrates “other” information provided to the hospital trauma team upon arrival at the hospital.

Counts/frequency: Age (39, 79.6%), Allergies (22, 44.9%), Assessment (24, 49.0%), Blood Gluclose (BG) (10, 20.4%), Cancer (0, 0.0%), Complaints (4, 8.2%), Diagnosis from OSH (4, 8.2%), eCO2 (2, 4.1%), Glascow Coma Scale (GCS) (7, 14.3%), Loss of Consciousness (18, 36.7%), Medications (18, 36.7%), Mental Status (12, 24.5%), Past Medical History (PMHx) (25, 51.0%), Pupil Dilation (4, 8.2%)

APPENDIX E

Clinical Emergency Procedure List

To be able to simulate real-world trauma transport, the team compiled a list of procedures that typically occur in an emergency setting as seen in Table 1. The set of procedures were determined by analyzing military tactical combat care guidelines and interviewing paramedics and trauma staff.

The following procedures were chosen for inclusion in simulations from a comprehensive list of pre-hospital trauma procedures for their 1) clinical importance and potential impact on clinical decisions by providers at the receiving facility, and 2) anticipated ability to discriminate and identify the procedure based on motion and/or heat map activity.

- Tourniquet application: When major bleeding is identified, bleeding is controlled using firm, steady pressure to the site using a tourniquet placed proximal to the injury. The tourniquet is placed around the injured extremity over a bony prominence if possible, and the tourniquet is tightened until bleeding stops.
- Application of combat gauze: In cases of significant bleeding, combat gauze can be applied by holding pressure to the site.
- Bag-valve-mask ventilation: Bag-valve-mask (BVM) ventilation includes application of a facemask with adequate seal to use an attached bag to provide ventilation. With one provider, the mask is held in place by one hand, usually the non-dominant hand, and the other hand squeezes the bag to generate airflow into the patient's lungs. Providers may switch hands to avoid fatigue, and in some situations two providers are used – one holds the mask in place while the other squeezes the bag.
- Oral airway**:** In patients who are obtunded and have no gag reflex, placement of an oral airway displaces the tongue anteriorly so that does not obstruct the airway and facilitates bag-valve-mask ventilation. Oral airway placement involves opening the mouth, displacing the tongue anteriorly, and inserting the oral airway into the mouth so that the distal tip is behind the base of the tongue.
- Endotracheal intubation: Among patients who are unable to breathe on their own, an endotracheal tube or supraglottic airway may be placed in the trachea or posterior pharynx, respectively, to provide oxygenation and ventilation. This procedure is generally performed from the head of the bed, although it can be performed from other approaches. Typically, the right hand is used to open the mouth, the left hand is used to insert laryngoscope and sweep the tongue to the left, and the right hand is used to introduce either the endotracheal tube or the supraglottic airway. The laryngoscope is then removed from the mouth and the endotracheal tube or supraglottic airway bulb secured in place by inflated a bulb using a syringe with the appropriate volume of air. Attaching the tube to a BVM then ventilates the patient and listening to breath sounds with a stethoscope and/or looking for color change on a colorimetric indicator that detected carbon dioxide confirm appropriate positioning.
- Cricothyrotomy: This procedure is performed when oxygenation and ventilation above the vocal chords is not possible. This procedure is generally performed

facing the patient's the side so that the non-dominant hand palpates and stabilizes the cricothyroid membrane, and the dominant hand is used to puncture the membrane and introduce the tube, which is then attached to a BVM for ventilation. A scalpel or hollow bore needle can be used to gain access to the trachea inferior to the vocal chords.

- Needle decompression and chest tube placement: These procedures are performed when a tension pneumothorax is suspected. For needle decompression, a large bore needle with a catheter is inserted into the pleural space, generally at the 2^{nd} or 3^{rd} intercostal space at the mid-clavicular line. A one-way valve is then applied so that air accumulating in the pleural space exits the thoracic cavity. For chest tube placement, the patient is positioned so that the arm on the affected side is above their head; a scalpel is used to make an incision in the mid-axillary line at the 3^{rd} or 4^{th} intercostal space, and blunt dissection is used to gain access to the pleural cavity above the rib. A clamped chest tube is then inserted into the thoracic cavity and directed posteriorly and superiorly, aiming for the apex of the thoracic cavity. This tube is then attached to suction.

- Chest compressions: Chest compressions performed as part of cardiopulmonary resuscitation (CPR) when patients become unresponsive, are not breathing, and do not have a pulse. Chest compressions are performed by using both hands, one on top of the other, to push the patient's chest a depth of at least 2 inches over the sternum between the nipples at a rate of ~100-120 compressions per minute. Chest compressions will be alternated with ventilation if there is only one provider available.

- Immobilization: Patients are immobilized to prevent additional injuries or exacerbate existing injuries such as spinal fractures. This is done by placing the patient on a spine board and placing a cervical collar around the cervical spine.

- Splinting: Obvious fractures or bony deformities are stabilized by securing the extremity to a splint using gauze or kerlex.

- Peripheral intravenous catheter placement: Placement of a peripheral intravenous (PIV) catheter is generally done in the patient's arm or leg to administer intravenous medications or fluid. Obtaining PIV access involves placement of a tourniquet proximal to the vein, palpation of the vein, cleaning the site, insertion of the catheter, holding pressure proximally, releasing the tourniquet, attaching the catheter to tubing or a leuer lock, generally with a twisting motion, and applying a dressing to hold the PIV in place.

- Intraosseus placement: Obtaining intraosseus (IO) access allows administration of medication into the bone marrow. IO placement may be performed on the sternum, proximal humerus, or proximal tibia, distal femur, or distal tibia. IO placement can be accomplished using a spinal needle that is screwed into the bone by hand or by IO kits that drill or inject IO needles into the bone. Once access to bone marrow has been obtained, the needle is secured in place with a dressing.

- Medication administration: Medications may be administered by PIV, intraosseus (IO) access in order to facilitate intubation, during a code, or to treat massive bleeding. Medications may be removed from a vial using a needle attached to a syringe, insertion of air, and then removal of the medication; this medication is administered after cleaning the hub of the PIV/IO, and in general 10ml of saline are used to flush the line. Medications such as succinylcholine, etomidate, and atropine may be stored in prefilled syringes, which can be attached directly to a PIV or IO.

APPENDIX F

EMG and Accelerometer Graphs Examples

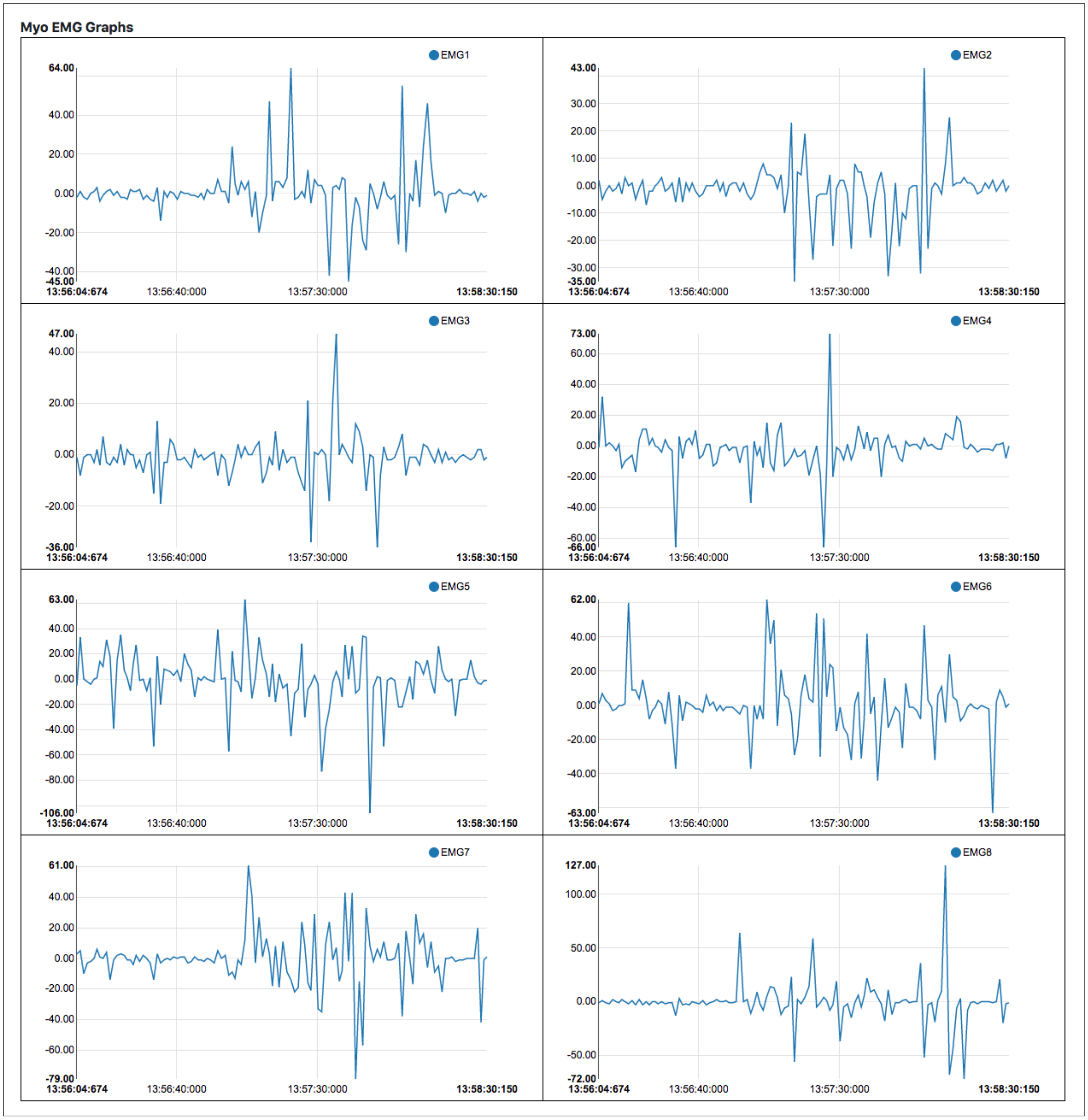


Figure 15: Myo activity during a simulation lab experiment.

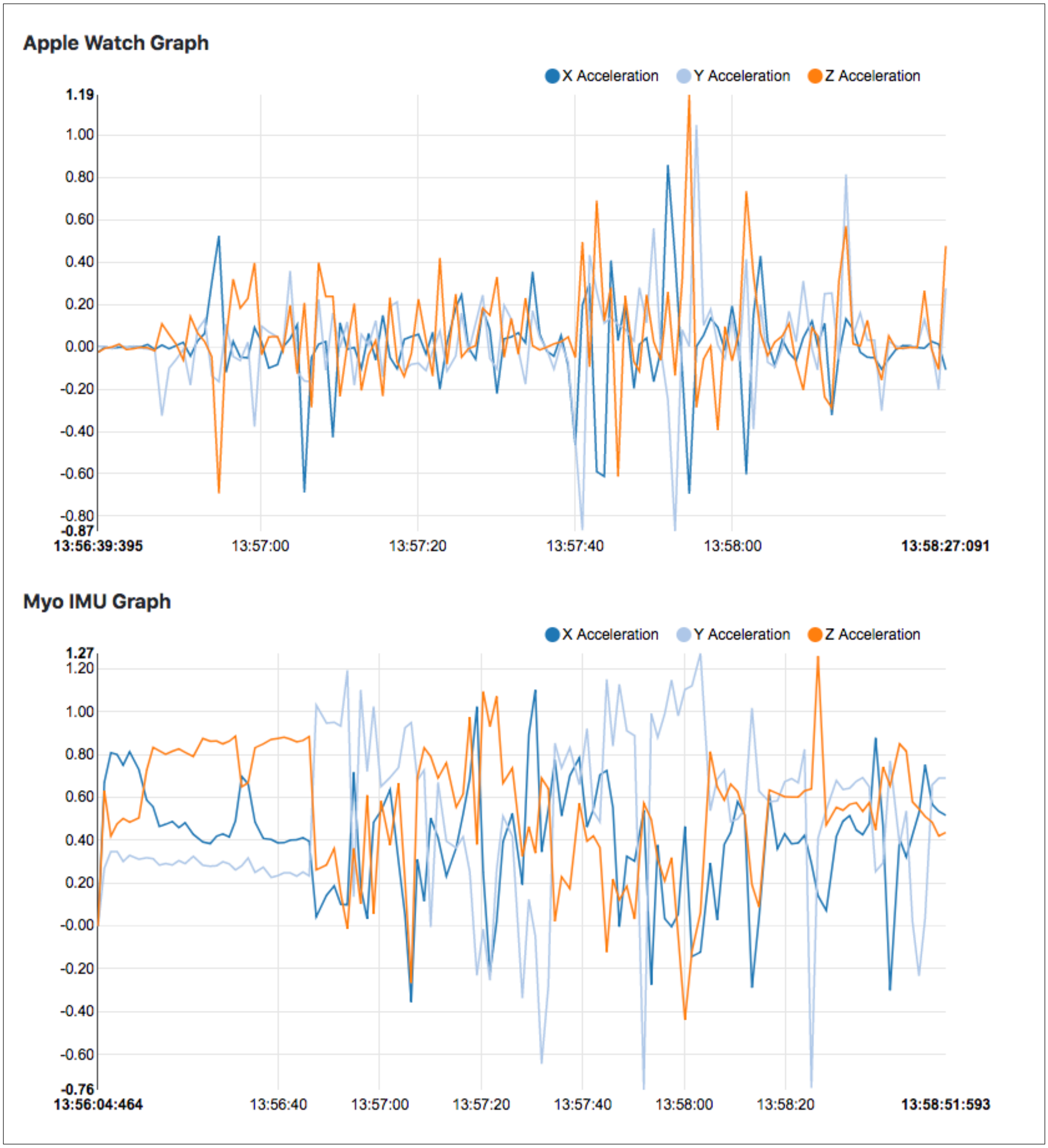


Figure 16: Myo Armband and Apple Watch accelerometer data. Myo does not remove gravity component so the two graphs are not directly comparable.

Event Displayed: CPR

Data	Description	ID	Start Time	End Time	Number of Data Points
Apple Watch Entire Session	Right CELA Test6	6B1CA978-5CCE-444C-ABDF-A9623B62A42D	may 23, 2018, 10:19:36 AM	may 23, 2018, 11:50:12 AM	20760
Apple Watch Plotted	Right CELA Test6	6B1CA978-5CCE-444C-ABDF-A9623B62A42D	may 23, 2018, 10:51:40 AM	may 23, 2018, 10:56:48 AM	346
Myo Entire Session	Right CELA Test6	201805231019004510771867301952688	may 23, 2018, 10:19:00 AM	may 23, 2018, 11:49:59 AM	15480
Myo Plotted	Right CELA Test6	201805231019004510771867301952688	may 23, 2018, 10:51:40 AM	may 23, 2018, 10:56:48 AM	258

Figure 17: Myo Armband and Apple Watch accelerometer data summary statistics for entire session compared to event specific CPR data, from one participant in a simulation lab experiment.

APPENDIX G

CELA Lab Setup, Process, and Data Collection:

The CELA lab served as the data collection environment and contained all the necessary equipment for the procedures in Appendix E. The repeated measures evaluation required each participant to complete each medical procedure multiple times within a three-hour timeframe. The target number of instances per medical procedure are provide in Table 4, where the desired number was determined based on pilot data. Bagging and CPR are time-based procedures; thus, the time per procedure was chosen such that fatigue effects may manifest. The procedures were grouped by category (i.e., airway management, wound related) and were completed in five rounds, where each round varied the presentation of the procedures to account for ordering effects.

Medical Procedure	Number of Instances	Medical Procedure	# of Instances
Administer IM Medication	5	Place an Oral Airway	10
Administer IO Medication	5	Place Blood-Pressure Cuff	5
Administer IV Medication	6	Place ECG Leads	5
Bagging	5, 7.5, and 10 minutes	Place an IO Line	5
Combat Gauze	6	Place a Pulse-Ox Monitor	5
Chest Decompression	5	Splinting	3
CPR	5 minutes	Intubation	2
King Airway	2	Combat Tourniquet	3

Table 4: Target Number of Instances per Medical Procedure

The evaluation's objective was to collect video, acceleration, and electromyography (EMG) data, which will serve as training and testing data for the medical activity detection system. Cameras were used to collect video data, while the Myo device and Apple watch collected acceleration and EMG data. Another evaluation goal was to have multiple participants with varying amounts of medical experience, as experience will impact the time to perform and the proficiency of each medical procedure.

One pilot and four participants completed the evaluation. The pilot was a study team member and served to motivate the experimental design and to validate the acceleration and EMG data collection systems. The four participants had varying amounts of experience (i.e., ranging from a medical student to emergency room surgeon) and consisted of one female and three males. Additionally, the amount of medical experience did not mean that the participant had more experience doing a procedure than another participant. For example, P3 had more years of experience than P2, but P2 had substantially more experience placing an IV. Future data collection will collect medical experience data by procedure, in addition to overall experience.

The task environment set-up can be seen in Figure 18, which depicts the placement of four cameras. Each participant was free to move around the bed to perform each medical procedure but were instructed to remain seated in a rolling chair. The necessary medical equipment was placed on the mannequin or on the bed, prior to completing the corresponding procedure.

Each medical procedure was tagged with start and stop times using the collected video. The average times (st. dev.) in seconds each participant took to complete each procedure are given in Table 5. Participant P2 tended to take less time administering medication through an IO line than the other participants; although, the participants took roughly the same time to administer medication through an IV line. Participants took roughly the same time to perform a round of CPR (breathing and compressions); however, participant P2's first CPR round contained 200 chest compressions. The reason for the larger number of chest compressions was that research has shown greater resuscitation success with the 200 initial chest compressions.

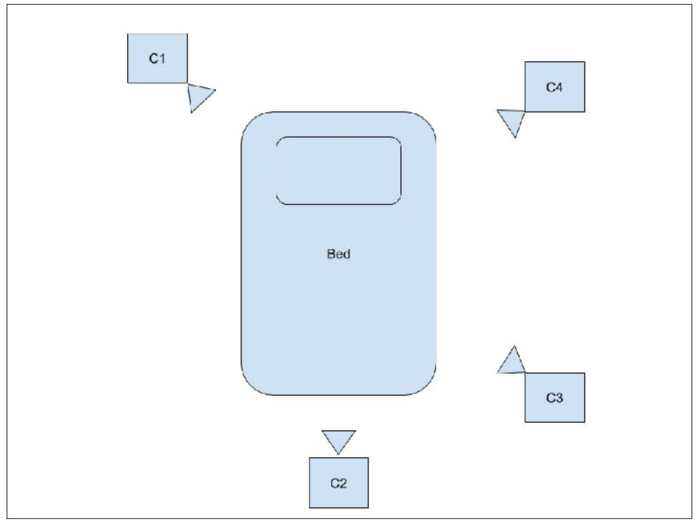


Figure 18: CELA Lab Environment Set-Up

There was a large variance in chest-tube timings, which may be attributed to the procedure's infrequency in the real-world. A large variance between participants was seen in the combat gauze timings, which was attributed to not having the necessary medical training equipment. Participant P4 took longer to draw medication than the other participants, due to the participant's relative inexperience. Participant P2 took less time than the other participants to put on the ECG leads, administer IM medication, place an IO line, place an IV line, tie an IV tourniquet, wrap a head wound, and place a pulse-ox monitor. Participant P2 had more experience doing these procedures than the other participants. The participants took roughly the same time to intubate a patient, but had some variance in inserting a king airway, which may be attribute to how much lubricant was in the dummies mouth and how much force was needed to insert the king airway. Participant P3 took a little longer to place an oral airway compared to the other participants, due to the use of a tongue compressor. Participants P3 and P4 took a longer amount of time to splint a leg than participants P1 and P2, due to wrapping additional gauze around the leg. Only participants P3 and P4 sutured the chest-tube incision close, and participant P4 took longer suturing, as this was the participant's first time placing a chest-tube. Additionally, participants P3 and P4 were the only participants to swab the chest-tube incision site with alcohol, which created longer event timings than participants P1 and P2. There was a large timing range in monitoring the patient's vital signs across participants, which may be due to the lack of vital signs, as the participants had to act like they were hearing breaths and heart-beats.

Overall, the medical event timings show the effect experience and individual differences have on completing a medical procedure. The medical activity detection system will have to account for various amounts of medical experience and times to achieve high accuracy. The system may be trained on certain movements, e.g., inserting a laryngoscope blade, rather than the entire medical procedure to improve classification performance. Additionally, the system may need to incorporate various window sizes (the length of time features are extracted from) in order to accommodate the range of timings across the medical procedures.

Procedure	P1	P2	P3	P4
Administer IO Med.	26.0 (13.89)	10.75 (3.4)	29.0 (6.56)	29.0 (0.00)
Administer IV Med.	28.5 (11.15)	24.5 (11.26)	20.6 (6.15)	24.0 (16.97)
Bagging	342.0 (217.79)	358.5 (263.13)	126.5 (144.23)	608.0 (0.00)
Blood-Pressure Cuff	39.5 (3.54)	11.33 (1.53)	20.0 (8.49)	15.0 (0.00)
CPR (Breath)	4.73 (1.16)	7.73 (1.35)	7.0 (1.26)	5.31 (0.63)
CPR (Compressions)	15.4 (0.74)	17.75 (1.96)	20.91 (22.59)	17.21 (3.45)
Chest-Tube	130.33 (22.23)	49.4 (26.37)	66.67 (23.71)	100.0 (0.00)
Combat Gauze	42.0 (31.11)	12.0 (3.61)	3.0 (0.00)	77.0 (0.00)
Combat Tourniquet	51.0 (0.00)	34.0 (9.64)	58.5 (0.71)	62.0 (0.00)
Draw Medication	12.5 (2.12)	9.5 (2.38)	9.15 (3.11)	22.8 (3.27)
ECG Leads	133.0 (4.24)	61.0 (10.44)	110.5 (2.12)	101.0 (0.00)
IM Administration	18.0 (14.14)	6.0 (2.16)	13.33 (6.02)	12.0 (2.83)
IO Line	60.0 (16.46)	33.25 (10.4)	78.67 (38.42)	36.0 (0.00)
IV Line	92.25 (33.03)	42.5 (11.81)	91.0 (40.8)	70.0 (1.41)
Intubation	39.0 (19.8)	42.0 (0.0)	43.0 (5.66)	39.5 (6.36)
King Airway	18.0 (8.49)	30.5 (12.02)	26.0 (1.41)	15.0 (0.00)
Oral Airway	4.29 (2.29)	3.88 (1.36)	7.0 (1.41)	5.5 (0.71)
Pulse-Ox Monitor	11.5 (3.54)	7.0 (1.73)	13.0 (2.83)	16.0 (0.00)
Splinting	46.0 (9.9)	46.0 (11.36)	61.0 (26.87)	65.0 (0.00)
Suturing	nan (nan)	nan (nan)	100.67 (25.01)	471.0 (0.00)
Swab Area w/ Alcohol	4.71 (1.25)	3.9 (1.91)	7.23 (3.65)	6.67 (3.14)
Tie Tourniquet (IV)	15.75 (3.2)	8.83 (3.49)	15.0 (5.77)	16.5 (4.95)
Vital Monitoring	16.5 (3.54)	22.33 (5.51)	17.67 (7.09)	27.0 (0.00)
Wrap Head Wound	51.0 (5.66)	33.33 (11.85)	60.5 (6.36)	68.0 (0.00)

Table 5: Medical Procedure Event Timing's Mean (Std.Dev) by Participant in Seconds

APPENDIX H

Hierarchal Task Analysis:

Each medical procedure was broken down into their anatomical movements using hierarchal task analysis [2] to identify distinct movements that may differentiate the procedure from other procedures, which will be useful for classification.[3] Published medical procedure guides were used to decompose each procedure into sub-tasks, which were broken into anatomical movements. The task analysis for CPR is given in Figures 18 and 19. The medical procedure (CPR) is decomposed into four sub-tasks, where subtask *1.3 Give 2 breaths* can be completed two different ways: without a bag-valve mask *(Subtask 1.3A)* or with bag-valve mask *(Subtask 1.3B). Subtask 1.3* is further decomposed into sub-sub tasks, which are decomposed into anatomical movements. This analysis allows for determining overlap between procedures and potential state changes. For example, there is overlap between CPR and Bagging, if CPR uses the bag-valve mask to give two breaths. However, the chest compressions in CPR are unique; thus, training a system to only use data pertaining to chest compressions to classify CPR may increase accuracy. A state change can also be identified between two subtasks, such as between *Give 2 Breaths* and *Chest Compressions*. Detecting state changes may be useful for state-based classification algorithms, such as Hidden Markov Models.

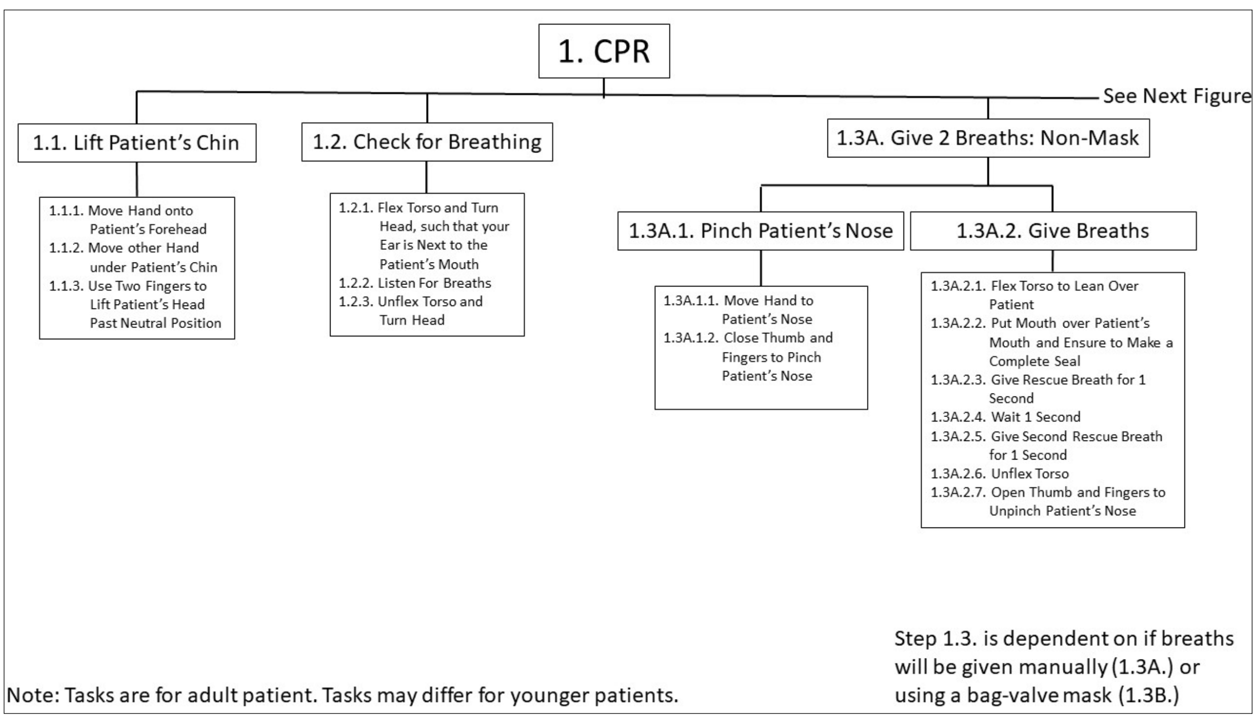


Figure 18: Hierarchical Task Analysis for CPR

[3] Stanton, N. A. (2006). Hierarchical task analysis: Developments, applications, and extensions. Applied ergonomics, 37(1), 55-79.

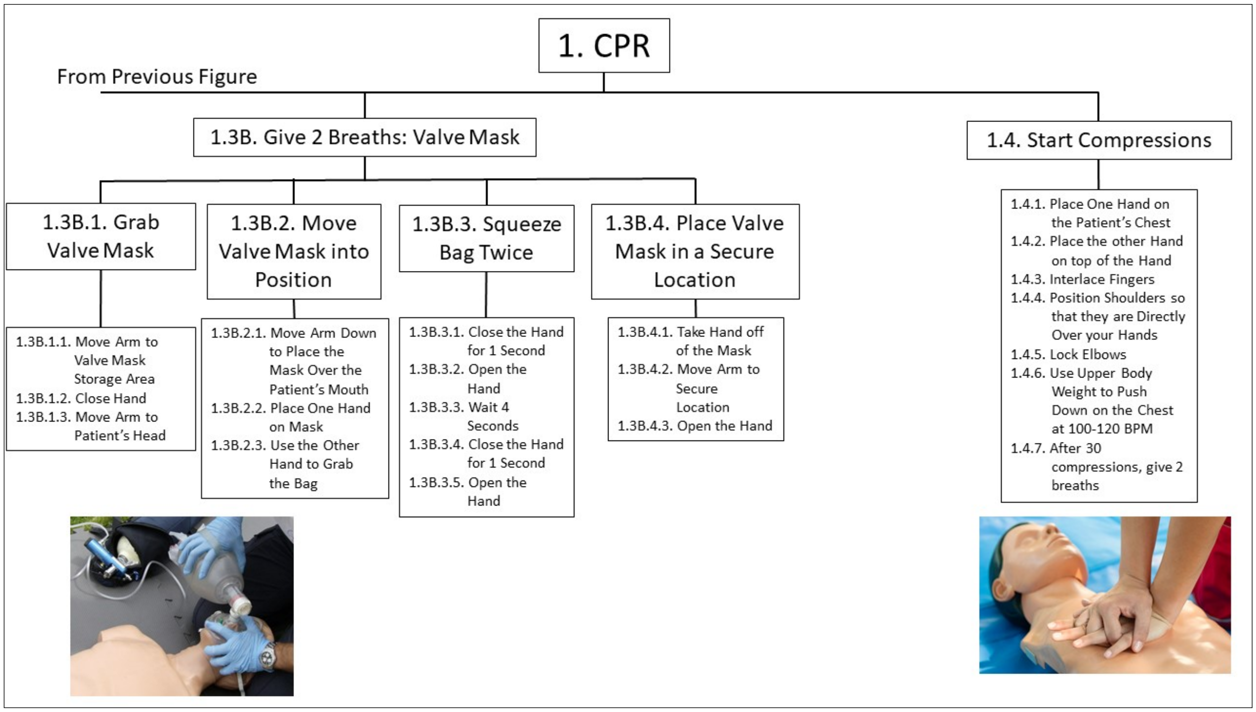


Figure 19: Hierarchical Task Analysis for CPR

APPENDIX I

Heatmap Generation for Emergency Medical Procedure Identification

Richard A. Paris[a], Paul Sullivan[a], Jamison Heard[a], Deirdre Scully[a], Candace McNaughton[b], Jesse M. Ehrenfeld[b], Julie A. Adams[c], Joseph Coco[b], Daniel Fabbri[b] and Bobby Bodenheimer[a]

[a]Vanderbilt University
[b]Vanderbilt University Medical Center
[c]Oregon State University

ABSTRACT

Ideal treatment of trauma, especially that which is sustained during military combat, requires rapid management to optimize patient outcomes. Medical transport teams 'scoop-and-run' to trauma centers to deliver the patient within the 'golden hour', which has been shown to reduce the likelihood of death. During transport, emergency medical technicians (EMTs) perform numerous procedures from tracheal intubation to CPR, sometimes documenting the procedure on a piece of tape on their leg, or not at all. Understandably, the EMT's focus on the patient precludes real-time documentation, however this focus limits the completeness and accuracy of information that can be provided to waiting trauma teams. Our aim is to supplement communication that occurs en route between point of injury and receiving facilities, by passively tracking and identifying the actions of EMTs as they care for patients during transport. The present work describes an initial effort to generate a coordinate system relative to patient's body and track an EMT's hands over the patient as procedures are performed. This 'patient space' coordinate system allows the system to identify which areas of the body were the focus of treatment (e.g., time spent over the chest may indicate CPR while time spent over the face may indicate intubation). Using this patient space and hand motion over time in the space, the system can produce heatmaps depicting the parts of the patient's body that are treated most. From these heatmaps and other inputs, the system attempts to construct a sequences of clinical procedures performed over time during transport.

1. DESCRIPTION OF PURPOSE

The purpose of this work is to automatically identify the clinical procedures EMTs perform during transport using off-the-shelf passive sensors such as video cameras and EMT-worn accelerometers (e.g., Apple Watch). A passive system, in which no active input is required, is necessary to avoid distracting the EMT away from patient care activities. Current documentation and communication to receiving medical teams includes hand-written notes and brief verbal reports, respectively. In both forms, the information presented to the receiving team can be incomplete and inaccurate. Supplementing these existing communication methods with an automatically produced list of clinical procedures with time stamps has the potential to more adequately prepare for the triage and downstream management of trauma cases.

2. METHODS

For this specific work, the system uses a single data source, video data feeds, to identify clinical procedures. The video feeds are processed with the computer vision system OpenPose,[1–3] which analyzes each frame to identify persons in the frame and identify their skeletons. The skeletons include 18 different key point positions including hands, feet and the head. These key points designate where in each frame the person and their extremities are. Given these key points, the system first identifies the patient using simple heuristics such as them being in the center of the frame and having minimal movement. Next, the system identifies the EMT as the person closest to the patient. Once the patient and the EMT are identified, the system constructs a 'patient space', which is a geometric space relative to the patient's body. The system then tracks the EMT's hands in the patient space (i.e., hands over the head or over the leg).

To simulate real-world trauma transport, the team compiled a list of procedures that typically occur in an emergency setting as seen in Table 1. The set of procedures were determined by analyzing military tactical

Table 1. List of procedures and number of times each subject was supposed to complete each procedure. *Indicates that this procedure took place on the left or right arm randomly.

Medical Procedure	Times Completed	Medical Procedure	Times Completed
Adminster IM Medication*	5	Place an Oral Airway	10
Adminster IO Medication*	5	Place Blood-Pressure Cuff	5
Adminster IV Medication*	6	Place ECG Leads	5
Bagging	3	Place IO Line	5
Combat Gauze on Arm*	3	Place Pulse-Ox Monitor	5
Combat Gauze on Head	3	Splint Arm*	3
Combat Gauze on Leg*	3	Splint Leg*	3
Perform Chest Decompression	5	Take out ETT Tube	2
Perform CPR	1	Take out King Airway	2
Perform Intubation	2	Take out Oral Airway	10
Perform King Airway	2	Tourniquet on Arm*	3
Place an IV Line (Left Arm)	3	Use Stethoscope to check vitals	5
Place an IV Line (Right Arm)	3		

combat care guidelines and interviewing paramedics and trauma staff. The list of procedures includes a span of procedure types including airway management, medication administration, and stabilization. Video of four subjects with various medical and emergency response training was then recorded of each subject performing the procedures in a simulation lab (Figure 1). Each of these subjects performed a number of iterations of each procedure. Repetition allowed for the detection of individual differences as well as repetitive differences.

The video data collection system was configured as follows. Video was recorded with four Apeman A20 4K action cameras, which record 3840 by 2160 pixels at 24 fps. Video data were collected from four angles for 3D reconstruction. The positioning of the cameras relative to the patient is shown in Figure 2. Each of these cameras are at a height of 2m to ensure that the patient and subject are visible in each camera. Camera 2 was selected so that the patients body would be centered in the frame and so that screen space would roughly correspond to a 2D plane directly over the body. The Apeman cameras generate a series of 181 second videos with one second of overlapping frames between clips in the series. The final one second was removed (24 frames) of overlapping video so that no duplicate processing is completed by OpenPose. Each 3 minute video was analyzed with OpenPose, which was running on an NVidia Docker virtual machine using two GeForce GTX Titan X GPUs.

Visual inspection by trained personnel in conjunction with specific *a priori* criteria (such as two fingers on the wrist) are used to determine the beginning and end points of a procedure. To determine the exact frame at which each procedure begins and ends, the trained personnel visually inspect the recording and tag frames. These beginning and end points are used to split the data into smaller procedure-specific chunks to be analyzed.

Given these procedure-specific video chunks, hand key points of the EMT are extracted and used to generate a Gaussian field around them. This extraction process is done for every frame and summed over all frames in the chunk. By summing intensities of the fields over all chunks and frames, a heatmap is generated over the body showing the most frequently occurring positions of the hands for the chunk (Figure 3). These heatmaps will be used as training data for a convolutional neural net classifier, which is intended to classify procedures.

3. RESULTS

Figure 3 shows the heatmaps generated from three procedures: intubation, insertion of an IV and splinting a leg. The background image represents the patient's body and the colors represent the position of the EMT's hands over the patient's body in patient space. The yellow color represents the areas above and around the patient

Figure 1. Still image taken from video data (left) and the same frame with OpenPose generated data overlaid to form a skeletal representation of both the patient and EMT (right).

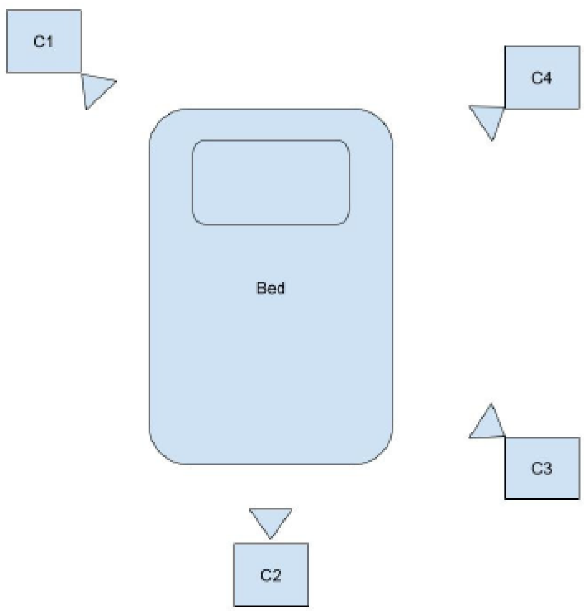


Figure 2. Positioning of the four cameras used during data collection. Each camera shot 4K video at 24 fps.

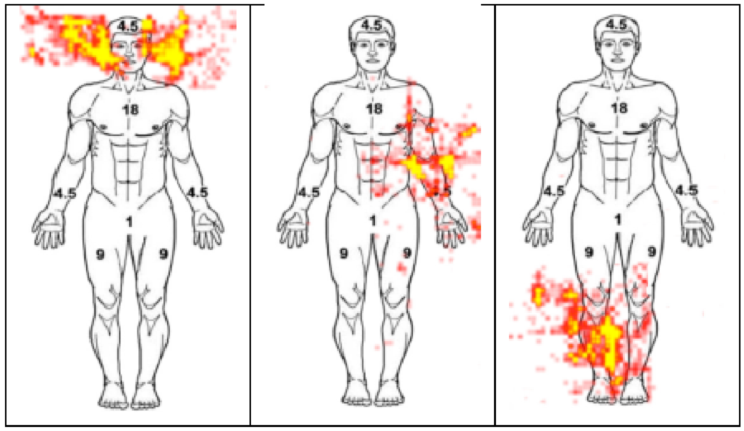


Figure 3. Heatmap showing the position of both the technician's hands over the patient's body during a single instance of intubation, insertion of an IV, and splinting of a leg (from left to right).

where the EMT hands are located most often. Visually these heatmaps indicate that we can identify the body part which is being worked on, which will help in determining which procedure is being performed.

4. CONCLUSIONS

This work presents a method to video record training data of medical procedures and visualize heatmaps of those procedures. This visualization allows inspection of a given set of procedures. These heatmaps can be used as training data to begin to classify procedures and may aid in computer identification of the procedure as it is being performed in emergency conditions. Since we intend to use this in conjunction with activity data gathered from other devices, this work shows a first step in how computer vision and machine learning can be used to help further identify the procedure being performed.

5. OTHER INFORMATION

This work is original and unpublished. It has not been submitted for publication or presentation in any form.

ACKNOWLEDGMENTS

This work was supported by the Department of Defense Contract Number W81XWH-17-C-0252 from the CDMRP Defense Medical Research and Development Program.

REFERENCES

[1] Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y., "Realtime multi-person 2d pose estimation using part affinity fields," in [*CVPR*], (2017).

[2] Simon, T., Joo, H., Matthews, I., and Sheikh, Y., "Hand keypoint detection in single images using multiview bootstrapping," in [*CVPR*], (2017).

[3] Wei, S.-E., Ramakrishna, V., Kanade, T., and Sheikh, Y., "Convolutional pose machines," in [*CVPR*], (2016).

APPENDIX J

Paper: Feasibility Assessment of a Pre-Hospital Automated Sensing Clinical Documentation System

Sean M. Bloos[1], Candace D. McNaughton, MD PhD MPH[2], Joseph R. Coco, MS[1], Laurie L. Novak, PhD MHSA[1], Julie A. Adams, PhD[5], Robert E. Bodenheimer, PhD MS[4], Jesse M. Ehrenfeld, MD MPH[1,3], Jamison R. Heard[4], Richard A. Paris[4], Christopher L. Simpson, MA[1], Deirdre M. Scully[4], and Daniel Fabbri, PhD[1]

[1]Department of Biomedical Informatics, [2]Department of Emergency Medicine, [3]Department of Anesthesiology
Vanderbilt University Medical Center, Nashville, TN
[4]Department of Electrical Engineering & Computer Science
Vanderbilt University, Nashville, TN
[5]School of Electrical Engineering & Computer Science
Oregon State University, Corvallis, OR

Abstract

Documentation in the pre-hospital setting is challenged by its limited resources and fast-paced, high-acuity nature. Military and civilian medics are responsible for performing procedures and treatments to stabilize the patient, while transporting the injured to a trauma facility. Upon arrival, medics typically give a verbal report from memory or informal source of documentation such as a glove or piece of tape. The development of an automated documentation system would increase the accuracy and amount of information that is relayed to the receiving physicians. In this paper, we discuss the 12-week deployment of an Automated Sensing Clinical Documentation (ASCD) system among Nashville Fire Department EMS paramedics. We examine the data collection methods, operational challenges, and perceptions surrounding real-life deployment of the system. Our preliminary results suggest that the ASCD system is feasible for use in the pre-hospital setting and revealed several barriers for which we found solutions.

Introduction

Military and civilian medics are responsible for retrieving, stabilizing, and transporting the wounded to a trauma facility[1]. However, accurately documenting medical care during transport is complicated for several reasons such as limited staff in the vehicle and care requirements for trauma patients[2]. Instead of documenting every activity during transport, medics typically give a brief verbal report to the receiving facility staff including chief complaint, mechanism of injury, vital signs and procedures performed[3]. This report may be supported by brief notes written on the patient, a scrap piece of paper, the medic's gloves, or in many cases relying only on the medic's memory. While the transmission of this information to hospital providers is essential for maintaining patient care and providing appropriate treatments, reporting is often incomplete[4,5] and likely inaccurate due to medic susceptibility to cognitive biases[3,6]. As a result, patient care may suffer[7].

Communicating some types of patient information can be done successfully with a verbal report (e.g., chief complaint, mechanism of injury, age, gender). However, specifics regarding the sequence of procedures performed, medication dosage and timing, and specific vital sign ranges are difficult to recall from memory given the high-intensity setting of trauma care. This information is essential for optimal care management, resource allocation and triage planning[8].

Our research objective is to develop an automated documentation system, which can detect pre-selected procedures performed by pre-hospital providers (i.e., paramedics or military field personnel) and create an abbreviated care record, without requiring the medic to actively produce the documentation. Instead of requiring active documentation, a range of sensors placed in the vehicle and worn on the medic that can passively aggregate data to describe care, from which algorithms can interpret and produce a care record. Candidate off-the-shelf sensors include electromyography (EMG) sensors, cameras, and inertial measurement units (IMUs), such as accelerometers.

This paper reports on the deployment of an automated sensing clinical documentation system. Specifically, we outline the equipment used, the configuration of the equipment in a civilian ambulance, perception of medics wearing devices, data collection processes and interfaces with the trauma facility. The system was deployed with the Nashville Fire Department (NFD) Emergency Medical Services (EMS) in conjunction with Vanderbilt University Medical Center, a level I trauma center in Nashville, Tennessee, which receives a high volume of acute trauma patients. Due to various concerns, we were not able to use cameras as a part of our system. While cameras are not a necessary component of the system, their inclusion may improve system accuracy. The lessons learned from this work will allow for the development of a more robust documentation system for medics and potentially improve patient outcomes.

Background

High quality healthcare requires effective and accurate communication among providers. The dynamic nature of care by medics in the pre-hospital setting can make it difficult to document procedures in real time and communicate vital clinical information to hospital providers. In their handoffs to emergency department (ED) staff, paramedics dedicate 75% of verbal reports to patient demographics and presenting signs/symptoms and only ~7% to pre-hospital treatments[9], even though pre-hospital treatments and clinical course largely drive resource allocation and treatments upon arrival[10]. To address the need for more efficient methods for accurate documentation and communication of pre-hospital care, we have developed a clinical documentation system, Automated Sensing for Clinical Documentation (ASCD). The system leverages a combination of off-the-shelf sensors to collect data from which algorithms attempt to detect the procedures that are preformed and create an abbreviated care record. This record is designed to be generated in real time, or near real time, and transmitted upstream to providers in the ED as a supplement to the verbal handoff. The goal of this technology is to increase the accuracy and detail of clinical information transmitted to upstream clinical providers and teams, particularly in high-acuity and trauma settings.

The current technology system was designed using feedback from pre-hospital personnel and hospital emergency department providers, and then was refined and validated in a controlled setting by performing procedures on simulated patients. In this study, we describe the implementation process, barriers encountered, and lessons learned during real-world, pre-hospital deployment of the technology.

Methods

The study protocols were reviewed and approved by the Institutional Review Board of Vanderbilt University. The Automatic Sensing for Clinical Documentation system was deployed among Nashville Fire Department (Nashville, TN) paramedics over a 12-week period. Nashville Fire Department provides fire protection and emergency medical care for 533 square miles and transports patients to numerous hospitals within the metropolitan Nashville area. In 2018, NFD responded to approximately 130,000 calls. Of note, NFD EMS protocols for airway management do not include the use of medication for rapid sequence intubation (RSI). Prior to deployment of the system, written informed consent was obtained from the participating paramedics.

Paramedic shifts were selected based on the availability of paramedics and research staff. At the beginning of each shift, a trained researcher, who is also a paramedic, equipped one paramedic with the ASCD system. In its original design, the system records video of patient care to track the medic's hand over the patient's body (e.g., around the patient's leg or chest) to infer the types of procedures that can possibly done at a given time (i.e., hands must be over the chest to do CPR). However, due to privacy concerns, only the motion sensing and EMG portion of the system, (i.e., Apple Watches, iPhones, and Myo Armbands) was deployed in the field setting for preliminary testing. In order to collect and transmit collected information, a laptop and other additional equipment was also carried by the research observer (see *Data Collection*).

For a portion of the participating paramedic's 12-hour shift, the trained research observer observed all clinical activity and recorded "start" and "stop" times for targeted procedures performed inside the ambulance (Table 1). These procedures were chosen based on focus groups with emergency medical service personnel and common procedures performed in the ambulance[11]. Procedures performed outside of the ambulance were not included in the recorded observations due to the distance from the laptop and Bluetooth receivers to the Myo Armbands and iPhones. The observer did not participate in any patient care.

For the subset of patients transported to Vanderbilt University Medical Center (VUMC), observations also included the handoff between the team of paramedics and the ED team; procedures and interventions performed during the ED visits were also documented. At the conclusion of each shift, the paramedic who wore the technology completed a debrief survey. These surveys featured a user-centered design approach, considered the context of use, specific requirements and areas of design optimization[12].

Between each observation, the ASCD equipment was cleaned using SaniWipes® and all components of the system were charged. No substantial damage was received to the equipment during use.

Table 1: Pre-hospital procedures assessed by the ASCD system and number of times observed			
Procedure	**Start**	**Stop**	**# times procedure observed during deployment**
Administer Intramuscular (IM) Medication	Medication taken out of the box	Needle withdrawn from patient	0
Administer Intraosseous (IO) Medication	Medication taken out of the box	Finished flushing with saline	0
Administer Intravenous (IV) Medication	Medication taken out of the box	Finished flushing with saline	3
Apply Pressure to Stop Bleeding	First applying pressure to artery	Hand leaves wound	0
Bagging	Touches bag	Hand leaves bag	0
Blood Pressure Measurement	Touches blood pressure cuff	Hands leaves blood pressure cuff	13
Chest-Tube Insertion	Touches scalpel	Apply tape/bandage to secure chest tube	0
Chest-Tube Preparation	Has hemostat in one hand and tube in the other	Hands leave hemostat	0
Chest-Tube Suturing	Touches needle-driver	Apply tape/bandage	0
Combat Gauze	Touches gauze	Patient's wrapped extremity is put down	0
Combat Tourniquet	Touches tourniquet	Writes time on tourniquet	0
CPR (Respiratory Support)	Touches patient's head	Hands leaves patient's head	0
CPR (Compressions)	Starts first compression	Hands leaves patient	0
Draw Medication	Removes syringe from packaging	Syringe leaves medication vial	1
ECG Lead Application	Touches ECG electrodes	Hands leave patient after final electrode applied	6
Intubation with Endotracheal (ET) Tube	Touches laryngoscope	Removes syringe after inflating the cuff of ET tube	0
Intraosseous (IO) Access	Touches IO drill	IO is secured to the patient	0
Intravenous (IV) Access	Touches IV from package	Taped IV down	11
Supraglottic Airway (SGA)	Touches SGA	Removes syringe after inflating the cuff	0
Oropharyngeal Airway (OPA)	Touches OPA	Hand leaves OPA after placement	0
Pulse Oximetry Monitoring	Touches pulse oximetry monitor	Hands leave pulse oximetry monitor	1
Splinting	Touches splint	Wrapped extremity is put down after splint is applied	0
Swab Area with Alcohol	Touches alcohol prep	Alcohol prep is out of user's hand	6
Swap Vial with Alcohol	Alcohol towelette is removed from packaging	Alcohol towelette is removed from vial	0
Tie Tourniquet (preparation for IV access)	Touches IV tourniquet	Hands leaves IV tourniquet	5
Vital Monitoring	Touches stethoscope or thermometer	Hands leave stethoscope or thermometer	5
Wound Dressing	Touches gauze package	Patient's wrapped extremity is put down	0

Data Collection

Transport of Equipment: The following equipment was housed in a Pelican® case during transport: 2 Myo® Armbands, 2 Apple Watches (Series 3), 2 Apple iPhone 7s, AT&T Unite Express 2 WiFi hotspot, 2013 Apple MacBook Air, and their accompanying chargers (Figure 1). The equipment was delivered at the beginning of each shift and kept with a member of the research team.

Figure 1: Pelican case setup

Electromyography: Thalmic Labs Myo® gesture control armbands were used as the source for EMG data collection. Myo armbands were connected to the MacBook Air using a Bluetooth connection. After the armbands were pulled onto the forearms, they were calibrated by having the paramedics position their hands as shown in Figure 2.

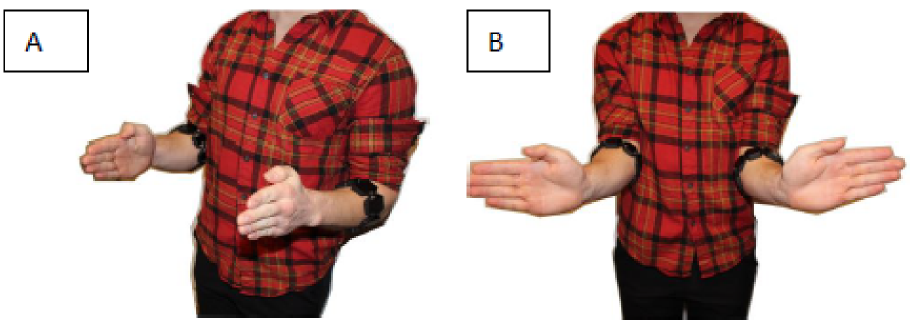


Figure 2: Myo neutral position (A) to Myo sync position (B) to calibrate the armbands

Accelerometer: One Apple Watch Series 3® was worn on each wrist with the watch-face outward and was paired with an Apple iPhone 7®. Both iPhones were stored in a secure cabinet in the rear of the ambulance. The iPhones were connected to a virtual private network (VPN), which was necessary to securely send data to a VUMC server. We originally used the PulseSecure application to access the VPN, but then the system migrated, causing us to switch to F5 Access. To establish internet connection, a AT&T Unite Express 2® WiFi hotspot was used.

Data Flow (Figure 3): Both the left and right Myo Armbands were connected to the laptop using a Bluetooth connection. The laptop was connected to the mobile hotspot over a VPN. The left and right Apple Watches were paired with a corresponding iPhone 7, which also used the hotspot WiFi.

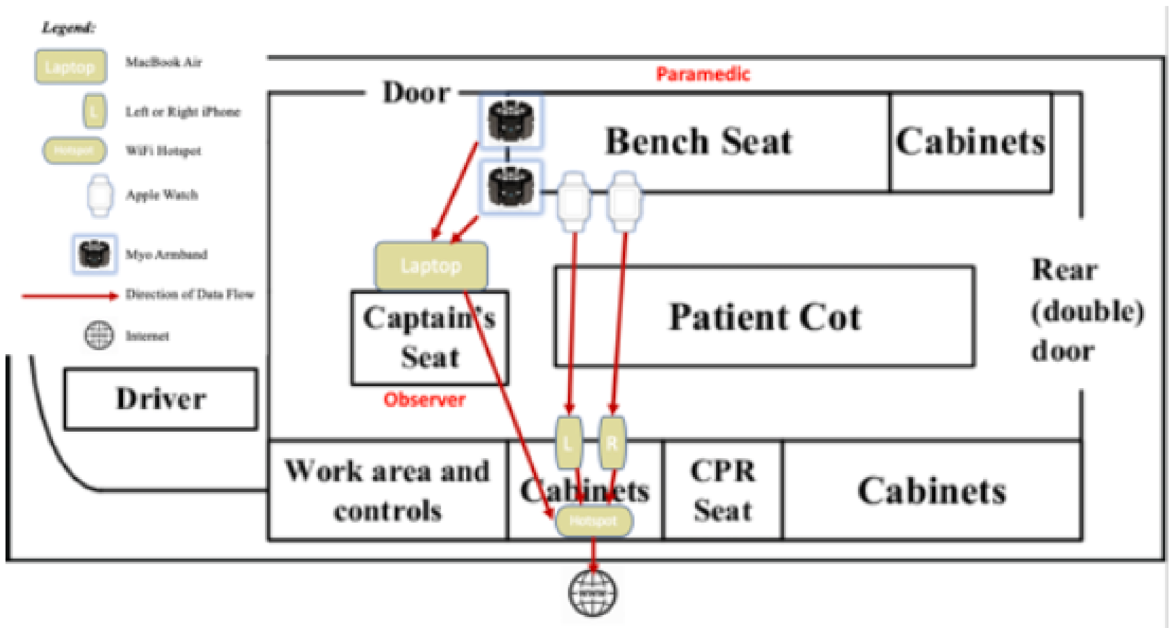


Figure 3: Flow of data inside the ambulance[13]

Data Transfer: At the conclusion of each shift, the Myo data collection files were transferred from the laptop to a shared cloud account (VUMC Box) for review by the data analysts. Due to the large file sizes, the transfer was completed using the VUMC WiFi instead of the hotspot.

Notification of VUMC research team: An application, Life360, was installed onto one of the iPhones. A geo-fence was set up around VUMC, which notified the clinical researcher when the phone was detected within 800 feet of the ambulance bay. In addition, the observer would send a message via Life360 to the clinical research when enroute to VUMC. The clinical researcher would meet the paramedic team at the ambulance bay and observe the handoff.

Documentation of Targeted Procedures:
The research team thoroughly debated the method for documenting the targeted procedures prior to deployment of the ASCD system. We needed a method that was quick, accurate and allowed for the research observer to note any discrepancies in how the paramedics actually started and stopped each procedure. Variations in procedural equipment or protocols could cause the start and stop times to differ. Original discussions included methods such as (i) keeping field notes and manually keeping track of time, or (ii) creating an iPhone application that allowed the observer to press a start/stop button with pre-selected procedures. The most significant challenge with the former was accuracy of procedure times. For the latter, we felt that an application with pre-selected options did not allow the observer much versatility to add other comments and raised concerns about accidental selection of a start or stop time. We ultimately used a simple Python logging application that allowed for free-text entries and recorded time stamps upon entry.

Research documentation began at the point in which the patient was loaded into the ambulance. During each transport, the observer typically sat in the captain's or "airway" seat (rear facing), located behind the head of the patient, and documented from that position. Once the patient was loaded into the ambulance, the observer would open the laptop, ensure proper connectivity of the system, and begin recording the procedures as they happened in real time. For example, if the paramedic was going to start an IV, the observer would type "IV start" into the log application when the paramedic touched the IV start kit. At the end of the procedure, indicated by the taping down of the IV, the observer would type "IV end". This denoted the gold standard for "start" and "stop" times for each procedure of interest (Table 2).

Table 2: Sample targeted procedure log.	
Procedure Occurred	**Procedure Description**
2019-01-09 12:42:13.030240	12 lead start
2019-01-09 12:42:38.984363	12 lead end
2019-01-09 12:43:07.886611	correction 3 lead
2019-01-09 12:44:53.242214	albuterol tx start
2019-01-09 12:45:11.890404	mask applied
2019-01-09 12:45:59.294573	tourniquet
2019-01-09 12:46:02.672012	iv start
2019-01-09 12:46:27.348737	IV in
2019-01-09 12:49:31.370297	IV procedure END
2019-01-09 12:49:54.562230	IV attempt fail
2019-01-09 12:57:29.503253	12 lead
2019-01-09 12:57:48.126144	12 lead end
2019-01-09 12:58:58.214071	albuterol tx end
2019-01-09 12:59:40.566805	check lung sounds

Paramedic Debriefs: After each observation, the research observer provided feedback to the research team regarding lessons learned in the field, barriers encountered, and feedback obtained. Survey responses from the paramedic participants were entered into REDCap, a secure web application designed for creating and managing online surveys and databases[14].

Results

Over seven observations, two paramedics wore the system for a total of 45 hours. We observed the transport of 16 patients to 6 different facilities and information after handoff was obtained for 6 patient encounters (Table 3). Using the first procedure logged as a start time and the end of the last procedure logged as an end time, we estimated the median time of active treatment during transport to be 8 minutes and 15 seconds with a standard deviation of 5 minutes and 24 seconds.

Table 3: Breakdown of patients transported to VUMC. (Psychiatric Assessment Service = PAS)			
Patient #	**ESI Score**	**ED Disposition**	**Chief Complaint / Mechanism of Injury**
1	2	Transfer to PAS	Suspected ingestion
2	2	Transfer to PAS	Overdose
3	2	Discharge	Auto vs. pedestrian
4	3	Unknown	Generalized weakness
5	2	Transfer to PAS	Suicidal ideations
6	3	Discharge	Intoxication/Chest Pain

Table 4: Challenges encountered during data collection and their solutions.	
Barriers	**Solutions**
Intermittent interruption in Myo Armband connectivity to laptop	Laptop location moved to the head of the stretcher, under the patient's head (Figure 4)
Script programs stopped recording when the laptop lid was shut	Installed disable lid sleep widget
Intermittent interruption in Apple Watch data collection	Implemented a live feedback system to visualize interruptions in data collection
Insufficient hotspot data for data collection	Data use was monitored proactively via web portal; ~1GB was needed per 6-hour observation
Confusing Apple Watch start/stop application	Start/stop feature changed from a tapping mechanism to a slide bar
Concerns of marrying pre-hospital observations to correct paramedic-to-ED handoff	Process developed to ensure consecutive subject data entry. Relative time (since start of paramedic shift) used to identify patients
Myo armbands intermittently vibrate if they are unsynced (caused by displacement of armband)	Paramedics were cautioned that this may occur, and they attempted to not desync the armbands. This vibrating functionality will be removed in future trials.
Systemwide VPN upgrade for VUMC users	We were forced to switch the VPN connection on the laptop. It had no apparent effect on data collection
Original hotspot data plan was canceled by the carrier for an unknown reason	Observations were delayed until we were able to obtain a new hotspot and data plan

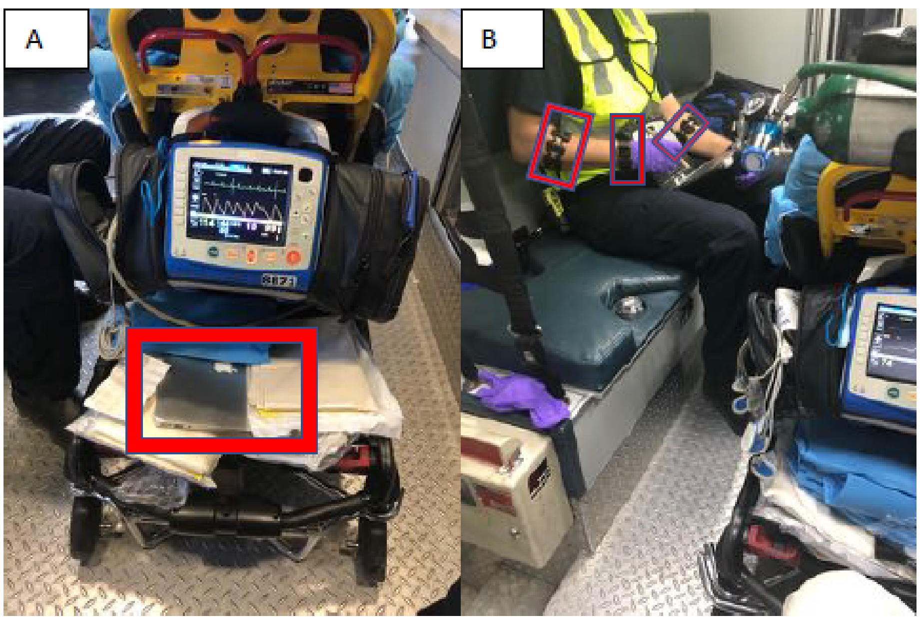


Figure 4: Placement of laptop in ambulance (A) and the ASCD system in use (B)

Current Documentation Techniques by Paramedic Participants

During the field observations, the majority of documentation occurred in the ambulance. For "non-critical" patients (i.e., stable vital signs), the paramedics typically used the charting software installed on their Toughbook® laptops to document items such as past medical history, current medications, drug allergies and demographic information. In addition, the cardiac monitors used by NFD had the ability to store vital signs such as blood pressure, heart rate, oxygen saturation and respiration rate. This log could then be uploaded directly to the patient care report following the call. During the care of patients who required more attention, or were more critical, documentation typically took place in the form of the paramedic writing on their glove. In other situations, the paramedics did not document some of the procedures at the time they were performed and simply documented them retrospectively from memory, which is considered standard care practice for EMS providers. Patient care documentation into the formal charting system typically occurred post-trip arrival. The paramedics suggested that each patient chart took approximately 30 minutes.

Operational Challenges

This study was set up based on the assumption that a large number of patients would be transported to VUMC. However, several factors indicated transport to other facilities. Those factors included whether or not VUMC was on diversion (i.e., not accepting patients), the hospital preference of the patient, proximity of the call to VUMC and the condition of the patient. Additionally, several scheduled observation periods were canceled due to mechanical malfunction of the ambulance, personnel illness, and work schedule changes.

Communication of Procedures to Receiving Facility Staff

We observed 6 handoffs between paramedics and ED clinical teams at VUMC. Two handoffs occurred at triage, where there was no physician present for the verbal report. Verbal reports from the paramedic to the ED staff largely consisted of the chief complaint and signs and symptoms. During the handoff of a level II trauma patient, the paramedic relayed all information regarding procedures (IV start, medication administration, vital signs, and cervical collar application) to the trauma staff.

Establishing Rapport with the Paramedic Participants

The research observer was a paramedic, and this allowed him to build a productive rapport with the paramedic team. The observer arrived at the beginning of each shift, which was at either 0530 or 1730 hours, depending if it was the night or day shift, respectively.

General Feedback from Paramedics

The paramedic participants indicated that they could wear the armbands for the duration of a 12-hour shift with no anticipated difficulties. Participants described the armbands as tight but reported that the bands did not restrict their overall movement or interfere with patient care. We noted that there was an impression on the paramedic's skin of the armbands that typically lasted for 30 minutes following their removal. Participants suggested considering other devices that might be more comfortable or fit in clothing. Other comments included that this technology would be useful in situations where there are critical or multiple patients.

Table 5: Results from paramedic questionnaires.

Factor	Response
Ability to wear entire shift	Yes – 7/7
Perceived comfortableness	Neutral – 1/7 Slightly Uncomfortable – 6/7
Likeliness to wear entire shift	Unlikely – 1/7 Likely – 2/7 Extremely Likely – 2/7 *Note: 5 responses due to addition of this question*
Issues with devices interfering with uniform	"They do not interfere with clothing" "No issues" "There are no complications with uniform and armbands"
Feelings regarding devices tracking movements	"I do not have any concerns" "I do not see any real problems with the devices tracking my movements"
Overall experience	"They get more uncomfortable the longer they are on. Even with all of the links taken out, they are a bit tight" "It has been a good experience"
Perceived feasibility of automated documentation	"I think it's completely feasible to have it automatically document time on action to improve documentation accuracy" "I feel like it would be very helpful in the pre-hospital setting with exact times and interventions"
Perceived usefulness of automated documentation	"It would be helpful on calls that require more hands on the patient where you don't have time to document as you go" "It would be useful when we are dealing with a critical patient or have multiple patients on the same scene. It would also be helpful to have this information to help give a report to the ED"

Methods for Future Data Analysis

This paper specifically does not discuss the algorithms used to convert the collected sensor data to an abbreviated care record. Briefly, the algorithms work as follows. First, the video data are used to track the position of the medic's hands over the patient's body using software such as OpenPose[15]. The medic's hand position acts as a prior function to determine the possible procedures that can be performed at a given point in time. Next, the other IMU and EMG data are summarized with various metrics (e.g., entropy, power, etc.). These data feeds are then fed into classifiers to predict what procedure is being performed, if any. Some data cleaning are also performed such as removing the gravity vector and removing vehicle vibration.

Discussion

To our knowledge, there have been no previous attempts to create and deploy such a sensor and documentation system using this range of sensors. As a result, we were able to identify challenges surrounding logistics, connectivity, evaluation, and perception by the paramedics.

The objective of the system is to support and supplement the documentation processes during transport (step B in Figure 5). For example, the ASCD system would be particularly useful in settings where verbal communication may be limited, such as the battlefield. The creation of an abbreviated care record with timestamps identifying which procedures were performed can enable upstream hospital providers to more effectively provide care.

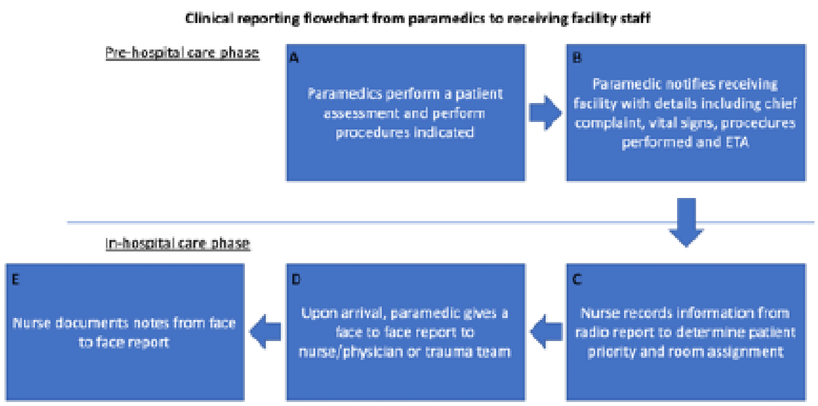


Figure 5: Clinical reporting flowchart from paramedics to receiving facility staff

We anticipated that deployment of the system into the field would induce more variation and therefore, subsequent error into our model. Detection of such error will allow us to refine the current algorithm used for the identification of targeted procedures. Additionally, deployment of the ASCD system will allow us to evaluate the perceived feasibility and usability of an automated system among paramedics in the pre-hospital setting.

Future processing will include using a deep neural network using both convolutional and recurrent layers with memory trained classifier in hopes for a greater efficacy[16]. We plan to move towards a real-time analysis using discrete, rolling windows for classification such that the start and stop of the event is not known by the classifier, which will be the case in real-world deployments.

Limitations

Our study was primarily limited by the number of paramedic participants (2). As we are still attempting to optimize the system and identify system failures, it was practical for a small-scale deployment. The reason for having a single observer and paramedic participant during the initial deployment was to identify these challenges before expanding the study. The second limitation can be attributed to the unknown nature of pre-hospital care. Paramedics are unsure of which procedures and treatment will be performed until they arrive on scene and assess the patient. As a result, there were some procedures for which we were not able to collect any data.

Conclusion

This paper reported on the lessons learned from deploying an automated sensing clinical documentation system in a real-world environment. It discussed challenges of configuring equipment, collecting data and dealing with failures. Many incremental steps were taken to reach the goal of a working system that could be safely deployed in the field and collect data, without interfering with care. Future work will analyze how well the algorithms are able to correctly identify which procedures are done during transport.

References

1. Sohn VY, Miller JP, Koeller CA, Rush, RM, From the combat medic to the forward surgical team: The Madigan model for improving trauma readiness of brigade combat teams fighting the global war on terror. J Surg Res. 2007;138:25-31.
2. Butler F. Tactical combat casualty care: combining good medicine with good tactics. J Trauma. 2003;54(5suppl):S2-3.
3. Carter AJ, Davis KA, Evans LV, Cone DC. Information loss in emergency medical services handover of trauma patients. Prehosp Emerg Care. 2009;13(3):280-5.
4. Sohn VY, Miller JP, Koeller CA, Rush, RM, From the combat medic to the forward surgical team: The Madigan model for improving trauma readiness of brigade combat teams fighting the global war on terror. J Surg Res. 2007;138:25-31.
5. Bergrath S, Skorning M, Rortgen D, Beckers S, Brokmann J, Mutscher C, Roissaint R. Is paper-based documentation in an emergency medical service adequate for retrospective scientific analysis? An evaluation of a physician-run service. Emerg Med J. 2011;28:320–4.
6. Staff T, Sovic S. A retrospective quality assessment of pre-hospital emergency medical documentation in motor vehicle accidents in south-eastern Norway. Scand J Trauma Resusc Emerg Med. 2011;19:1–11.
7. Laudermilch DJ, Schiff MA, Nathens AB, Rosengart MR. Lack of emergency medical services documentation is associated with poor patient outcomes: a validation of audit filters for prehospital trauma care. J Am Coll Surg. 2009;210(2):220-7.
8. Therien SP, Nesbitt ME, Duran-Stanton AM, Gerhardt RT. Prehospital medical documentation in the Joint Theater Trauma Registry: a retrospective study. J Trauma Acute Care Surg. 2011;71(1):103–8.
9. Evans SM, Murray A, Patrick I. Assessing clinical handover between paramedics and the trauma team. Injury. 2010; 41:460–4.
10. King DR, Initial care of the severely injured patient. N Engl J Med. 2019; 380:763-770.
11. Carlson JN, Karns CN, Mann C, Jacobson KE, Dai M, Colleran C, & Wang HE. Procedures performed by emergency medical services in the United States. Prehospital Emergency Care. 2016;20:1,15-21.
12. Preece J, Rogers Y, Sharp H, editors. Interaction design: beyond human-computer interaction. New York: John Wiley; 2002.
13. Kibira D, Lee TY, Marshall J, Feeney AB, Avery L, Jacobs A. Simulation-based design concept evaluation for ambulance patient compartments. Simulation. 2015;8:691-714.
14. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377.
15. Cao Z, Simo T, Wei SE, Sheikh Y, Realtime multi-person 2D pose estimation using part affinity fields. Conference on Computer Vision and Pattern Recognition. 2017:7291-99.
16. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui G, Corrado G, Thrun S, Dean J. A guide to deep learning in healthcare. Nature Medicine. 2019;25:24-9.

APPENDIX K

Paper: Understanding the Information Needs and Context of Trauma Handoffs to Design Automated Sensing Clinical Documentation Technologies: Qualitative Mixed-Method

Understanding the Information Needs and Context of Trauma Handoffs to Design Automated Sensing Clinical Documentation Technologies

Laurie L. Novak, PhD MHSA[1]; Christopher L. Simpson, MA[1]; Joseph Coco, MS[1]; Candace D. McNaughton, MD PhD[2,4]; Jesse M. Ehrenfeld, MD MPH[3]; Sean M. Bloos[1], Daniel Fabbri, PhD[1]

[1]Department of Biomedical Informatics, [2]Department of Emergency Medicine, [3]Department of Anesthesiology

Vanderbilt University Medical Center, Nashville, TN

[4]Geriatric Research Education and Clinical Center, Tennessee Valley Healthcare System VA Medical Center,

Abstract

Background: Current methods of communication between the point of injury and receiving medical facilities rely on verbal communication, supported by brief notes and the memory of the field medic. This communication can be made more complete and reliable with technologies that automatically document the actions of field medics. However, designing state-of-the-art technology for military field personnel and civilian first-responders is challenging due to the barriers researchers face in accessing the environment, and understanding situated actions and cognitive models employed in the field.

Objective: To identify design insights for an automated sensing clinical documentation (ASCD) system, focus. We sought to understand what information is transferred in trauma cases between pre-hospital and hospital personnel, and what contextual factors influence the collection, management, and handover of information in trauma cases, in both military and civilian cases.

Methods: Using a multi-method approach including video review and focus groups, we developed an understanding of the information needs of trauma handoffs and the context of field documentation to inform the design of an automated sensing documentation system that uses wearables, cameras and environmental sensors to passively infer clinical activity and automatically produce documentation.

Results: Comparing military and civilian trauma documentation and handoff, we found similarities in the types of data collected and the prioritization of information. We found that military environments involved many more contextual factors having implications for design, such as the physical environment (heat, lack of lighting, lack of power) and the potential for active combat and triage creating additional complexity.

Conclusions: Ineffectiveness of communication is evident in both the civilian and military worlds. We used multiple methods of inquiry to study the information needs of trauma care and handoff, and the context of medical work in the field. Our findings informed the design and evaluation of an automated documentation tool. The data illustrated the need for more accurate recordkeeping, specifically temporal aspects, during transportation, and characterized the environment in which field testing of the developed tool will take place. Employing a systems perspective in this project produced design insights that our team would not have identified otherwise. These insights created exciting and interesting challenges for the technical team to resolve.

Introduction

When military personnel or civilians are injured, field medics are the first to respond. Their objectives include stabilizing and transporting the patient to a trauma facility. Optimizing patient outcomes depends on accurate information sharing between field personnel and receiving physicians, including the context of the injury and clinical interventions. [1] These patient and information transfers in combat settings are highly variable and can range from minimal communication, e.g., pointing to a limb with a tourniquet when a helicopter picks up a patient from the scene in hostile territory, to verbal handoff when the patient is transported directly to the next higher level of care. When written documentation is generated, the documentation process may distract vital cognitive efforts away from patient care. Moreover, both written and verbal communication methods are vulnerable to rapid changes in clinical status, human cognitive biases, and mistakes in data collection, processing, and sharing. As a result, the information may be incomplete, inaccurate, or lost in communication [1–3]. Multiple handoffs further complicate the process and likely increase the risk of errors and miscommunication during transport.

Timely and accurate clinical documentation occurs when a sociotechnical system is designed and optimized around the relevant people, tasks, technologies, and physical and social environments [4]. Challenges include time pressure, the unique stress of providing care in combat situations, the use of personal protective equipment, limited visibility, and constrained working spaces. Additionally, even when documentation is generated, it is rarely transmitted in a way that is timely, clear, or effective [5]. Given the challenges of using traditional technologies to document clinical care at the point-of-injury and during transport, new systems are needed that can ensure better, more consistent, and clear communication among care teams.

Our main research effort is to develop novel technologies to automatically generate a clinical care record without requiring the active participation of personnel in the field. This automated sensing clinical documentation (ASCD) technology observes the tasks the medic performs using a combination of sensors. During its observation, the system outputs the list of clinical procedures that are being performed, ideally with high accuracy.

Designing ASCD involves understanding the other elements of the sociotechnical system into which the ASCD must fit. These elements include information the system must capture and the social and physical contexts in which it will be deployed. Direct assessment of the current state of military trauma handoffs is impractical due to safety and logistical concerns[6]. Relying on civilian ambulance observations produces data from a limited number of trauma cases, typically in an environment that is unlike a military field operation. Therefore, through a multi-modal analysis including focus groups and trauma-bay video review, this paper analyzes current trauma handoff practices to categorize information needs and contextual factors involved in trauma handoffs.

Background

The overall objective of our project is to develop an ASCD system that can be used on the battlefield by military personnel or by civilian medics in the field. The technology will involve a combination of off-the-shelf sensors, accelerometers, and cameras aligned with a software system that automatically detects the motion signatures associated with key clinical tasks and generates an abbreviated care record, which can be transmitted upstream in real-time. The system will passively collect data from a combination of accelerometers and cameras. Machine learning, activity detection and summarization algorithms will analyze the collected data to construct an abbreviated care record. This care record will provide patient clinical status, interventions, and anticipated resources needed upon arrival, without requiring active input from personnel in the field. Open research challenges to building such documentation systems include accuracy of predicting clinical events, usability, and deployment robustness.

In the United States' conflicts in Iraq and Afghanistan, the nation has suffered total deaths of 4,432 and 2,351, respectively as of December 4, 2019, [7] Since many fatalities occur between the point of injury (POI) and the medical treatment facility (MTF), the military has incorporated the use of Tactical Combat Casualty Care (TCCC) cards to document mechanism of injury, injury locations, vital signs and symptoms, and treatments .[8–10] This allows the first responders to triage the most critical patients in the pre-hospital (e.g. battlefield, vehicle) environment. [9,10] The military's documentation of the treatment during this period "is critical to ensuring continuity of care."[11] After completing the card, the first responder attaches the TCCC Card to the patient in a visible location as the record of treatment provided. Medical personnel in the receiving MTF are instructed to include the TCCC Card with the paper medical record, and enter the TCCC data into the patient's electronic health record (EHR) and appropriate trauma registry. Despite some evidence of a lack of compliance with the policy, the Defense Health Agency states, "The military will continue to use the TCCC Card until it fields a pre-hospital documentation platform that supports an electronic version."[11]

The transfer of a patient from a field medic to a MTF is a handover, defined as "the transfer of professional responsibility and accountability for some or all aspects of care for a patient, or group of patients, to another person or professional group on a temporary or permanent basis"[12] Handovers in health care have received significant attention in recent years as a period of high risk for the patient's safety. A review found that, in handovers between medics and hospital-based emergency departments, the key issues were: lack of common understanding, lack of active listening, variable quality and quantity of information exchanged, lack of clear leadership, lack of teamwork skills, busy and complex environment, and repetition of handover.[13] Organizations have tried to resolve issues with handovers through interventions to standardize communications, with mixed results.[14,15] Our project uses a systems perspective to examine an under-studied topic that is especially challenging in military medicine: the capture of clinical documentation in the field, especially in battle conditions.

Findings in this paper are organized with a health care systems engineering model that has been extensively used in the study of both handovers [16] and information technologies [17,18]. The Systems Engineering Initiative for Patient Safety (SEIPS) is a systems approach for understanding human activity in its context [19]. The fields of human factors and industrial engineering spurred the development of the framework to help frame research and innovation as technology was introduced into all areas of health care. The model was subsequently extended as SEIPS 2.0 to incorporate patient engagement, patient work, and work practice adaptations [20].

Methods

The research questions guiding this work were:

1. What information is transferred in trauma cases between pre-hospital and hospital personnel?
2. What contextual factors influence the collection, management, and handover of information in trauma cases?

Methods included 1) structured review of routinely captured videos of trauma handoffs in the Vanderbilt University Medical Center (VUMC) Emergency Department (ED), and 2) focus groups with ED providers, pre-hospital personnel such as emergency medical technicians (EMT) and paramedics, and military field medics. The research was conducted at VUMC and the Army's Rascon School of Combat Medicine on Fort Campbell, KY.

The study protocols were reviewed and approved by the Vanderbilt University Institutional Review Board. Given the infeasibility of observational research of the activities of front-line military medical personnel, we used triangulation of data [21] from two different methods to gather information about the work of field medics, a.k.a. pre-hospital personnel, and the handoffs between pre-hospital and hospital personnel.

Research Site

Vanderbilt University Hospital provides trauma care for 65,000 square miles. The Division of Trauma at Vanderbilt University Hospital handles close to 5,800 acute traumas admitting 4,300 of those annually. Essential for the quality of trauma care provided by Vanderbilt University Hospital are its facilities. These include a 20-bed burn unit, a 31-bed integrated Acute and Sub Acute care unit, which contains a 14 bed ICU, a 7 bed Acute Admission Area and a 10 bed Sub-Acute unit, and LifeFlight, an active air medical transport program. The Trauma Units' unique geography allows close integration and management of patient progress from admission to discharge. LifeFlight provides rapid access to the tertiary care facilities of the Trauma Center for all patients within a 140-mile radius of Nashville. In addition to LifeFlight, Vanderbilt receives patient transport from local and rural Emergency Medical Services (EMS).

Trauma video reviews:

VUMC Level I trauma cases are recorded for quality improvement purposes and reviewed weekly. These videos capture the pre-brief (in which EMS personnel and trauma team members from the ED and trauma team review facts about the arriving case and discuss a plan of action) and treatment while in the ED trauma bay. We reviewed fifty randomly selected videos to identify information transmitted via conversations during the handoff from EMS to hospital personnel. Videos are stored with no identifying linkages to patients and are deleted after a specified period. The videos were a way for us to observe handoffs without any patient-identifying information being collected.

A structured form facilitated the collection of relevant data from the videos. In order to refine a preliminary data collection form for the reviews, five videos were reviewed and documented by three reviewers. After the videos were reviewed, discussion of the results and any discrepancies in documentation were moderated by an independent arbiter. The reviewers came to a consensus on the types of information transferred from pre-hospital to hospital personnel and developed a data collection form to be used by a single, expert observer. The observer, a registered nurse, has extensive experience in trauma nursing and experience with review of the handoff videos. This observer viewed 50 trauma handoff videos, recording observations on the forms. After completion of the reviews, the data from the observation forms were entered into a REDCap[22] database for analysis and tabulation.

Focus Groups:

We conducted four focus groups. Two included civilian pre-hospital personnel (ambulance-based medics and aircraft-based flight medics), one included hospital personnel (physicians) and one was conducted with military personnel who provide medical care in the field.

The goals of the focus groups were to gather information from providers and medics with trauma experience to: 1) identify information transmitted in handoffs; 2) identify gaps in current handoff procedures, and 2) understand the social and physical context into which the technology will be deployed. The sessions explored participant experience transporting patients to the hospital, including elicitation of actual experiences in a combat environment when possible. Questions posed during the focus groups included:

- What information is normally shared during handoffs?
- What information is most useful to determine next steps in care management?
- Why/how is this information shared?
- What information is not useful to determine care management?

Based on the information shared in the session, we added probes to better understand the physical actions involved in transporting patients from the field/scene to the hospital including the implications of incorporating wearable technologies, cameras, and other devices into the process.

The sessions were audio recorded and transcribed for analysis. The transcripts were analyzed using a qualitative data analysis tool, Dedoose™. Given the variety of information shared by participants on information needs and context, we used an open coding procedure, identifying all themes that arose in the data. Three researchers coded the data, supported by discussion in frequent team meetings about findings and organization of the data. We then organized the data using the SEIPS 2.0 model for presentation and consideration by the team's technology designers.

Findings

Trauma video reviews

The handoff videos revealed information that is routinely relayed to the hospital team from the pre-hospital team. Figures below describe the content of each category of information in the 50 videos. Categories of information included: clarifying questions asked by the receiving medical team, procedures performed, mechanism of injury, medications and fluids given during transport, time of intervention/injury, blood pressure, heart rate, respiratory rate, oxygen saturation, and episodes of hypotension changes in clinical status.

Upon analysis of the data, it became apparent that clarifying questions were an important part of the pre-hospital to trauma team handoff. Clarifying questions are defined as questions from the hospital team directed to the pre-hospital team during handoff that are intended to obtain additional information that was not provided in the initial handoff. Of the 50 videos reviewed, 40 (80%) contained clarifying questions.

The clarifying questions that we observed in the videos consisted of questions about medication(s) (dosages, timing, etc.), personal medical history (if known), Glasgow Coma Scale or other (mostly neurological) exam results, time and mechanism of injury, allergies, whether or not restraints were used in accidents in vehicles, length of time tourniquet(s) have been in place, and fluctuations in vitals or neurological signs (blood pressure, heart rate, respiratory rate, oxygen saturation, etc.).

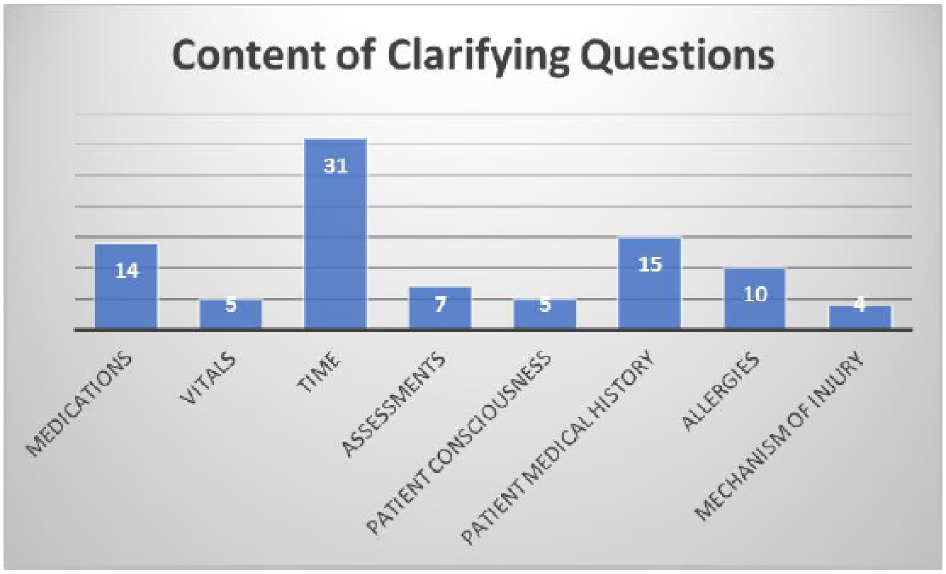


Figure 1. Frequency of clarifying question topics in the trauma videos.

The results for the other categories of information captured during observations are detailed in the following figures:

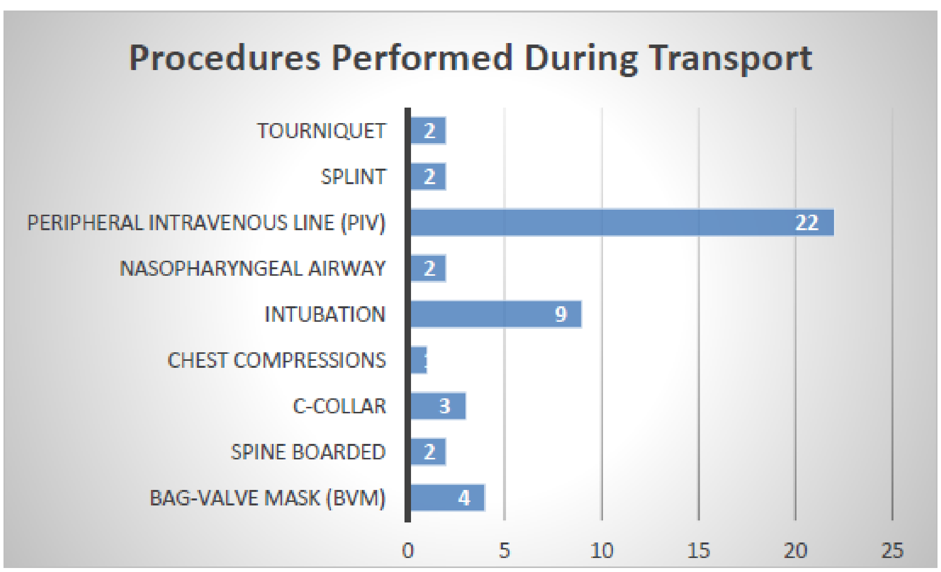


Figure 2. Frequency of procedures performed during transport in the trauma videos.

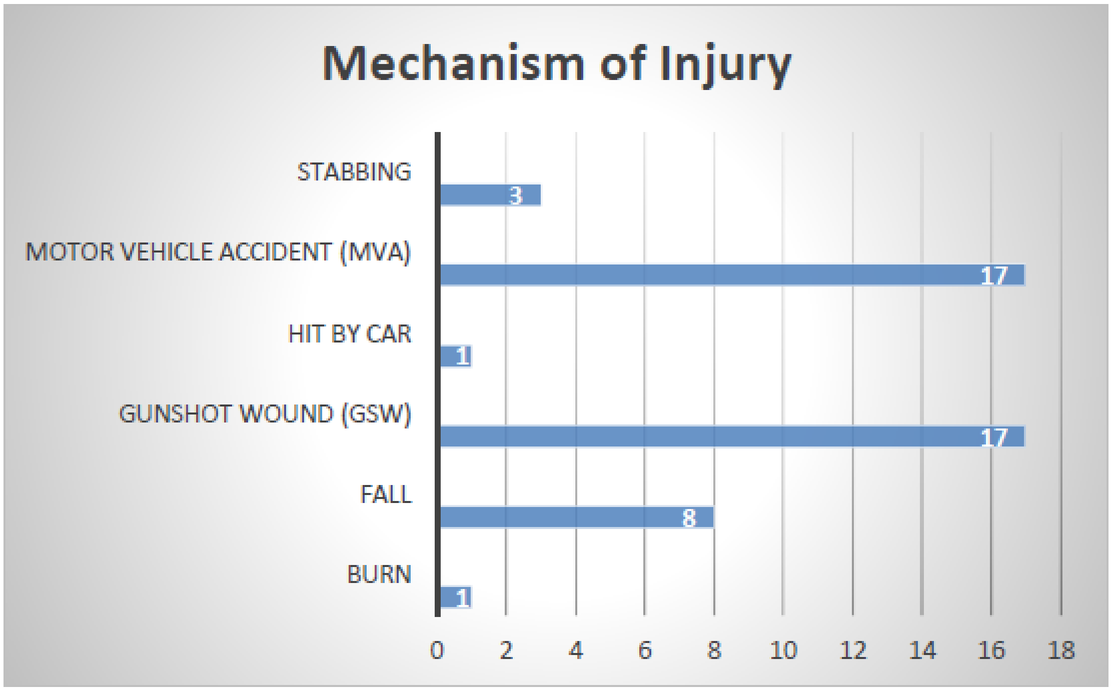


Figure 3. Frequency of each mechanism of injury in the trauma videos.

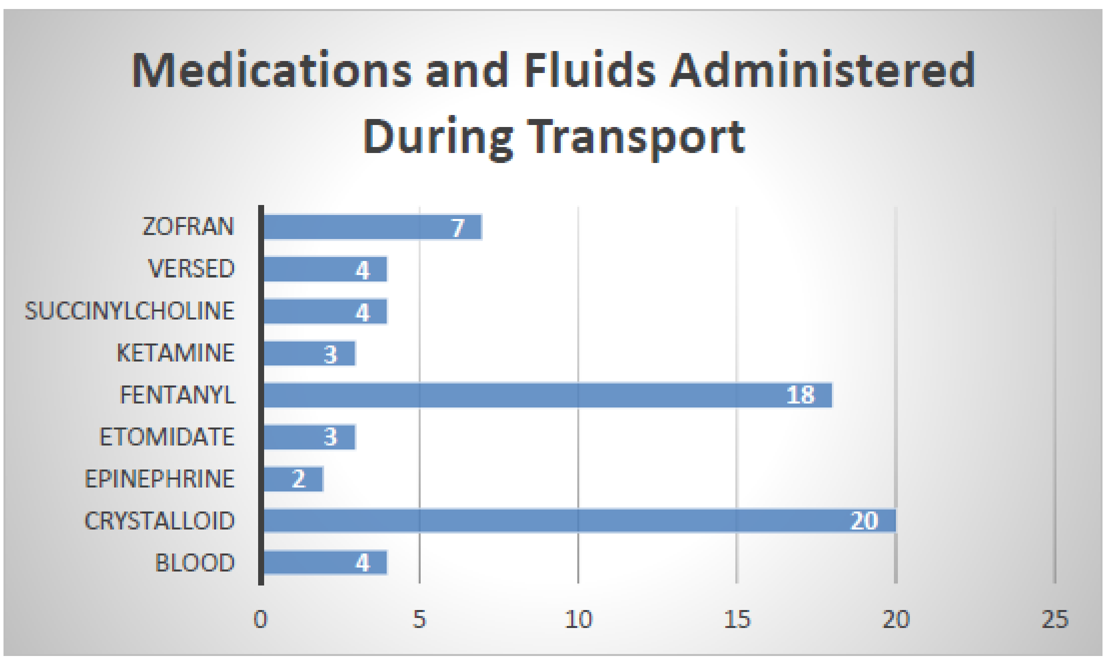


Figure 4. Medications and fluids administered during transport in the trauma videos.

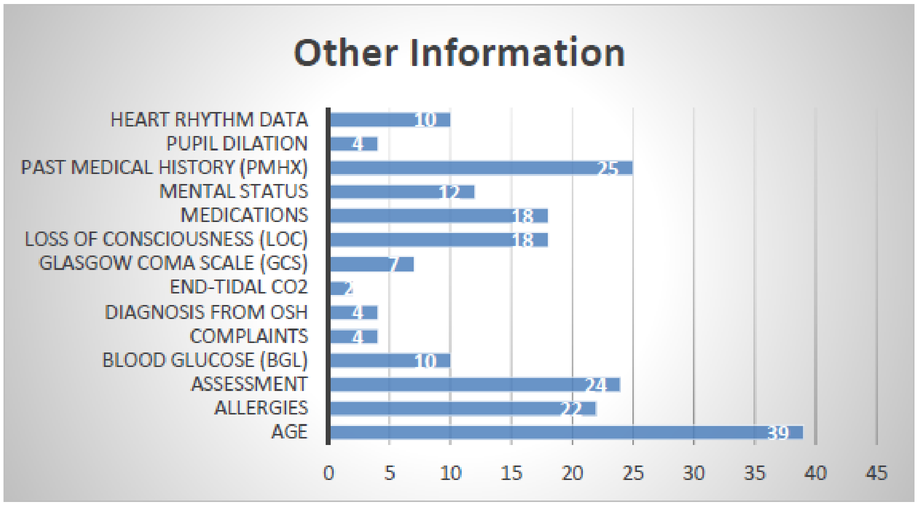


Figure 5. Other handoff information reported in the trauma videos.

<u>Focus Groups</u>

We conducted four focus groups comprising 19 participants. Participants included pre-hospital personnel (ambulance-based medics and aircraft-based flight medics), hospital personnel (physicians) and military medical personnel. Findings are summarized using the SEIPS framework on Table 1

Table 1. Work System Analysis for documentation in the field

	Civilian Pre-hospital System	Military Pre-Hospital System	Insights for the development of hardware and software tools
Technology and tools	- Written or electronic documentation of pre-hospital care - Gloves, paper, tape (for recording information) - Monitor	- TCCC (universal documentation card) o Sometimes partially completed by servicemember prior to mission - Communication headsets - Medics carry medical gear <u>and</u> combat gear	- Ad hoc methods are used, determined by environment - Information transmitted in advance can help hospital allocate resources - Simple statistic representing level of medic activity could give early indication of severity of patient injury - Mounting a camera in the vehicle is a challenge due to privacy issues - Object detection algorithms could potentially detect specific medication packages
Tasks	-Information captured o Vital signs o Demographics o Medications o Allergies o Time of events o Procedures o Pain level o Mechanism of injury -Procedures -Documentation	-Information captured o Vital signs o Procedures o Mechanism of injury -Procedures -Documentation -Triage -Active battle tasks -Radio communication	Priority information for handoff - Timing and sequence of procedures can suggest cause and effect - Worst and most recent vital signs are most useful
Organization	Information systems in EMS vehicle did not communicate with hospital	Large-scale, contracted military technology implementations sometimes lack coordination in	Transmitting information to hospital can reduce miscommunication, but also result in information overload

	Emergency Department	technology updates, resulting in lost communication between system components.	
Physical environment	Extreme heat is common	- Often extreme heat, exacerbated by excessive gear - Dusty - Noise level high in all settings - Taking notes is difficult - Rough terrain/ vehicle unstable - Low light in combat settings	- Need lightweight, small sensors. Armbands will be hot and uncomfortable - Voice technology not feasible because of noise - Sensors should conserve power when not in use - Wearable devices must withstand substantial amount of sweat from wearer
External environment		Mass casualty situations result in minimal documentation.	

Discussion

Findings from the videos illustrated that the most medically important information is not always effectively conveyed during the handoff from pre-hospital to hospital personnel. Of note were the clarifying questions observed during the review of the videos of the handoffs. Clarifying questions were observed in 80% of recorded handoffs, and most commonly involved temporal aspects of the case. Temporal questions included queries about the time the injury occurred, when a procedure was performed, and when a medication was given. Temporally-based questions were present in 27 of the 40 videos in which clarifying questions were asked of the pre-hospital staff. The next most commonly asked clarifying question involved either medications (drugs given, doses, timing, etc.) or the patient's past medical history. Both types of questions were present in 13 out of the 40 videos in which clarifying questions were asked during the handoff.

Data from the observations supports the findings from the three focus groups that more accurate information is needed at the time of handoff, specifically regarding time and sequences of procedures and/or medications. The hospital focus group emphasized that the most important information needed by the trauma team involved timing of events, especially regarding sequence of procedures performed during transport. The trauma videos revealed mechanisms of injury that would be less common in military environments, e.g. falls and being hit by a motor vehicle. However, we note that it is difficult to speculate on what types of trauma injuries may be seen in future combat situations, and it is likely short-sighted to design only for wounds produced by gunshots or explosions.

The pre-hospital and hospital teams have different priorities and/or capabilities in the performance of their roles in their respective environments. Pre-hospital teams need to get the patient in the vehicle and perform needed procedures during transport so that they can get the patient to superior resourced care teams, which is usually surgical intervention. Meanwhile, the receiving trauma team wants to be able to appropriately allocate resources based on procedures performed and patient trajectory during transport. These differences result in an inadvertent conflict about the priority of recording specific times of medication administration and/or performance and sequence of procedures during transport.

The findings from the video review and focus groups produced insights that informed device choices, software development, and evaluation strategy. Some surveillance technologies such as microphones that could potentially be useful to support documentation are not practical for noisy and insecure military field settings. While no tool will be able to capture every aspect of pre-hospital care, documentation through automated sensing can potentially enable medics to offer a more complete handoff to the receiving hospital.

<u>Implications for design</u>

Various activities are detectable through sensors. We identified numerous opportunities to capture activity (such as medical procedures or administration of medications) through motion detection and the relationship of motion signatures to locations on the patient's body, and the use of physical artifacts such as medication packaging. However there is heterogeneity in how procedures are performed and noise in the data. A robust system of data collection and analysis will be needed to deal with the forces of real-world deployments.

Challenges such as vehicle motion and sensor failure due to environment (e.g. a wearable sensor exposed to extensive sweat) may be universal. Challenges specific to military environments include the lack of lighting, high possibility of network failure, and possibility of active battle conditions while treatment is being carried out.

Conclusion

Ineffectiveness of communication is evident in both the civilian and military worlds. We used multiple methods of inquiry to study the information needs of trauma care and handoff, and the context of medical work in the field. Our findings informed the design and evaluation of an automated documentation tool. The data illustrated the need for more accurate recordkeeping, specifically temporal aspects, during transportation, and characterized the environment in which field testing of the developed tool will take place. Employing a systems perspective in this project produced design insights that our team would not have identified otherwise. These insights created exciting and interesting challenges for the technical team to resolve.

References

1. Carter AJ, Davis KA, Evans LV, Cone DC. Information loss in emergency medical services handover of trauma patients. Prehosp Emerg Care 2009;13(3):280–285.

2. Walderhaug S, Meland PH, Mikalsen M, Sagen T, Brevik JI. Evacuation support system for improved medical documentation and information flow in the field. Int J Med Inf 2008;77(2):137–151.

3. Scott LA, Brice JH, Baker CC, Shen P. An Analysis of Paramedic Verbal Reports to Physicians in the Emergency Department Trauma Room. Prehosp Emerg Care 2003;7(2):247–251.

4. Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inf 1999 Aug;55(2):87–101.

5. Therien SP, Nesbitt ME, Duran-Stanton AM, Gerhardt RT. Prehospital medical documentation in the Joint Theater Trauma Registry: a retrospective study. J Trauma Acute Care Surg 2011;71(1):S103–S108.

6. King DR. Initial Care of the Severely Injured Patient. N Engl J Med 2019;380(8):763–770.

7. Casualty Status, Department of Defense [Internet]. [cited 2019 Mar 11]. Available from: https://dod.defense.gov/News/Casualty-Status

8. Headquarters D of the A. Casualty Care, HQDA ATP (Army Techniques Publication) 4-02.5. Department of Defense; 2013.

9. U.S. Army Combined Arms Center. Tactical Combat Casualty Care, No. 17-13 Handbook v5. Department of Defense; 2017.

10. Joint Chiefs of Staff. Health Service Support, Joint Publication 4-02. Department of Defense; 2012.

11. Defense Health Agency. Implementation Guidance for the Utilization of DD Form 1380, Tactical Combat Casualty Care (TCCC) Card, Defense Health Agency Procedural Instruction #6040.01. 2017.

12. British Medical Association. Safe handover: safe patients. Guid Clin Handover Clin Manag Lond BMA 2004;7:141.

13. Bost N, Crilly J, Patterson E, Chaboyer W. Clinical handover of patients arriving by ambulance to a hospital emergency department: a qualitative study. Int Emerg Nurs 2012;20(3):133–141.

14. Cohen MD, Hilligoss PB. The published literature on handoffs in hospitals: deficiencies identified in an extensive review. BMJ Qual Saf 2010;19(6):493–497.

15. Talbot R, Bleetman A. Retention of information by emergency department staff at ambulance handover: do standardised approaches work? Emerg Med J 2007;24(8):539–542.

16. Agha RA. Handover in trauma and orthopaedic surgery–a human factors assessment. Ann Med Surg 2012;1:25–29.

17. Carayon P, Cartmill R, Blosky MA, Brown R, Hackenberg M, Hoonakker P, Hundt AS, Norfolk E, Wetterneck TB, Walker JM. ICU nurses' acceptance of electronic health records. J Am Med Inform Assoc [Internet] 2011 Jun 22; [doi: 10.1136/amiajnl-2010-000018]

18. Martinez VI, Marquard JL, Saver B, Garber L, Preusse P. Consumer health informatics interventions must support user workflows, be easy-to-use, and improve cognition: applying the SEIPS 2.0 model to evaluate patients' and clinicians' experiences with the CONDUIT-HID intervention. Int J Human–Computer Interact 2017;33(4):333–343.

19. Carayon P, Hundt AS, Karsh BT, Gurses AP, Alvarado CJ, Smith M, Brennan PF. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006;15(suppl 1):i50–i58.

20. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, Rivera-Rodriguez AJ. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669–1686.

21. Denzin NK, Lincoln YS, editors. Collecting and Interpreting Qualitative Materials. Thousand Oaks: Sage Publications; 1998.

22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377.

APPENDIX L

Paper: Automatic Clinical Procedure Detection for Emergency Services

Jamison Heard[1], Richard A. Paris[1], Deirdre Scully[1], Candace McNaughton[2], Jesse M. Ehrenfeld[2], Joseph Coco[2], Daniel Fabbri[2], Bobby Bodenheimer[1], *Senior Member, IEEE*, and Julie A. Adams[3], *Senior Member, IEEE*

***Abstract*— Understanding a patient's state is critical to providing optimal care. However, information loss occurs during patient hand-offs (e.g., emergency services (EMS) transferring patient care to a receiving hospital), which hinders care quality. Augmenting the information flow from an EMS vehicle to a receiving hospital may reduce information loss and improve patient outcomes. Such augmentation requires a noninvasive system that can automatically recognize clinical procedures being performed and send near real-time information to a receiving hospital. An automatic clinical procedure detection system that uses wearable sensors, video, and machine-learning to recognize clinical procedures within a controlled environment is presented. The system demonstrated how contextual information and a majority vote method can substantially improve procedure recognition accuracy. Future work concerning computer vision techniques and deep learning are discussed.**

I. INTRODUCTION

Communicating patient information accurately is vital to improving patient outcomes, but this information is typically not fully communicated from emergency services (EMS) to the receiving hospital [1]. This miscommunication is attributed to over or under-triaging the patient's state, resulting in incorrect trauma bay activation and a reduction in patient outcomes [2]. A noninvasive system that detects clinical procedures automatically can augment the current EMS communication flow in order to better alert receiving hospitals of the patient's triage level and reduce mortality rates. Such a system can draw from human activity recognition algorithms in order to accurately recognize clinical procedures and send procedural data, without medic input.

Human activity recognition is used to identify human activities in real-world scenarios [3] by relying on wearable or external sensors to collect activity specific patterns. Wearable sensors are physically attached to a human in order to collect movement and physiological data, while external sensors (i.e., cameras) are noninvasive and rely on voluntary human interaction. Features (or activity specific patterns) are extracted from the sensor data and are used by machine-learning algorithms to infer the current activity.

A human activity recognition algorithm has been shown to detect Cardio-Pulmonary Resuscitation (CPR) accurately using video data [4], but other commonly performed procedures indicative of trauma have not received any attention. This work attempts to recognize twenty-three clinical procedures using wearable sensors and video data. Further, a generalizable framework for documenting medical activity is defined. The wearable sensors capture a medic's arm movements and muscle contractions, but the data is insufficient to classify such a wide range of procedures. Video data is used to localize the medic's hand positions, relative to a patient, in order to determine an active body region or on which body part the medic is performing a procedure. Determining the active body region culls the number of potential procedures to recognize, as certain procedures are only performed in specific body regions (i.e., placing an oral airway only occurs near the patient's head). This class set reduction improves clinical procedure recognition accuracy; however, additional improvements are needed in order to realize a real-world automatic clinical procedure system.

II. EXPERIMENTAL DESIGN

The Center for Experiential Learning and Assessment lab at Vanderbilt University served as the data collection environment and contained the necessary clinical procedure equipment. The repeated measures evaluation required each participant to complete each procedure multiple times within a three-hour timeframe. The procedure list is provided in Table I and was chosen based on focus groups with emergency services personnel, army combat care guidelines, and commonly performed procedures in ambulances [5], [6], [7]. Four participants (one female and three males) with varying levels of medical training (i.e., a medical student, an emergency room surgeon) completed the evaluation.

Certain procedures were broken into sub-procedures in order to reduce overlap between procedures and body regions. CPR was decomposed into chest-compressions (Compressions) and giving the patient breaths (Breath), as the sub-procedures are performed on separate body parts. Swab area with alcohol was separated from multiple procedures (e.g., Intravenous Therapy (IV) and Intraosseous Infusion (IO) line), due to the overlap between procedures.

The task environment consisted of four cameras placed around a gurney, in which was an adult medical mannequin. Each participant was free to move around the gurney when performing each procedure, but were instructed to remain seated when possible in a rolling chair, as EMS personnel typically perform procedures while seated. The necessary medical equipment was placed on the mannequin or on the

This work was supported by the Department of Defense Contract Number W81XWH-17-C-0252 from the CDMRP Defense Medical Research and Development Program.

[1]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, Tennessee

[2]Vanderbilt University Medical Center, Nashville, Tennessee

[3]Collaborative Robotics and Intelligent Systems Institute, Oregon State University, Corvallis, Oregon

gurney, prior to completing the corresponding procedure. Each procedure video was tagged with start and stop times.

III. Clinical Procedure Detection System

The clinical procedure detection system combines wearable sensors with vision-based localization in order to accurately detect the medical procedures in Table I. The wearable sensor data captures arm movements that are representative of a procedure; however, there is a vast array of clinical procedures that need to be detected, which increases the problem's complexity. This complexity is reduced by determining the "active body region" using image processing.

A. Wearable Sensor Data Processing

The Myo device [8] is worn on each of the participant's forearm and captures arm movements and muscle contractions via an inertial measurement unit (IMU) and an 8-channel electromyography (EMG) sensor, respectively. Acceleration and orientation data is captured at 50 Hz, while the EMG data is captured at 200 Hz. The Myo automatically calculates the IMU's roll, pitch, and yaw. A 5 second window, with a 1 second stride, is applied to each sensor signal. Various window sizes were analyzed, but the 5 second window produced the best results.

The signal's mean, standard deviation, and max value are calculated for each window and are typical features extracted for activity recognition [3]. Each sensor signal is transformed into the frequency domain using the fast fourier transform in order to calculate the signal's spectral entropy. Thus, four features are extracted from each sensor signal resulting in fifty-six features per medic hand.

B. Image-Based Hand Localization

An orthogonal approach to classification using wearable sensor data is to use image processing to track the medic's hands during the clinical procedures. Many procedures are localized to certain areas on a patients body, making relative hand location a enticing factor. The image-based hand localization system determines the patient's closest limb to the medic's hands for a particular procedure and uses that information for classifier refinement.

OpenPose [9] is an image-based human body pose detection framework that generates 18 skeletal keypoints using the COCO system in screen space pixel coordinates for both the medic and the patient. The OpenPose parameters were tuned to accommodate a prone individual. An example output is provided in Figure 1. This image data is pre-processed to ensure consistency across each frame by ignoring frames when two bodies, (a medic and a patient), are not identified. The patient body is assumed to be the body whose centroid is closest to the center of the screen, due to the camera angles.

During a procedure, assuming the medics hands are proximal to the patient eliminates the need for 2D to 3D image conversion. Thus, the calculated distance between the medic's hand keypoints and each skeleton keypoint on the patient is in pixel space. This measurement's variability and noise is reduced by averaging the limb position over 1 second (24 frames) in order to determine the patient's closest limb to the medic's hands per second. The closest limb is mapped to one of four body regions: head, chest, arm, or leg.

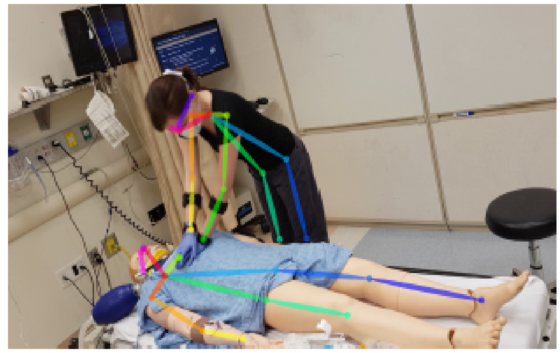


Fig. 1. OpenPose Output during CPR.

C. Clinical Procedure Classification

The extracted features from the Myos' IMU and EMG sensors are fed into a random forest classifier, which is a supervisory-based machine-learning algorithm that is an ensemble of individually trained decision tree classifiers. The random forest classifies a signal by taking the class mode of the decision tree ensemble. 100 decision trees with a max-depth of 500 are used for this work, where the parameters were chosen based on classifier performance.

The targeted domain requires knowing if a procedure was performed, not that every single window is correctly classified. Assuming a procedure's start and stop time is known, the procedure can be classified as the majority vote of each classified window within the procedure time frame. For example, if CPR (Compressions) consists of fifteen windows where ten windows are classified correctly and the other five windows are not, then the procedure can be correctly classified as CPR. Algorithm 1 provides the pseudo code for this classification. The algorithm cycles through each window between the procedure start and stop time, extracting

Algorithm 1 Clinical Procedure Classification Algorithm

Input: Procedure Start/Stop Time, Wearable Sensor Data, Video Data
Output: ProcedureClassification
```
PredictedProcedureList = []
for each window between Procedure Start and End time do
    Features = ExtractFeatures(window, WearableSensorData)
    ActiveBodyRegion = DetermineBodyRegion(window, Video Data)
    Classifier = DetermineClassifier(ActiveBodyRegion)
    Procedure = Classifier.Predict(Features)
    PredictedProcedureList.append(Procedure)
end for
return Max(ProcedureCount(PredictedProcedureList))
```

features from the wearable sensor data for each window. *DetermineBodyRegion()* runs OpenPose on the window's image data and determines the window's active body region, which is used to determine which trained random forest classifier to apply. The extracted features are fed into the classifier to predict a clinical procedure for the window. After each window is processed, the algorithm returns the Majority Vote of the predicted procedures using *Max(ProcedureCount())*.

D. System Validation

The clinical procedure detection system is validated using leave-one-subject-out cross-validation, where the random forest classifier is trained on two participants' randomly shuffled data and tested on the third participant's data. Participant Two's data was not analyzed, due to a camera failure during data collection. Five random forest classifiers were trained per cross-validation fold. One classifier was trained using data from every clinical procedure, which represents not knowing the active body region. The other four classifiers correspond to a body region (i.e., head, chest, arm, or leg) and were trained using the respective procedure data. The collected dataset created a class imbalance between procedures, which decreases performance. Thus, the over-represented procedures are randomly down-sampled during training in order to better balance the class set.

The cross-validation analysis was applied to three conditions: Unknown Body Region, Perfect Body Region, and Estimated Body Region. The unknown body region condition allows for analyzing how the clinical procedure detection system performs without image data (i.e., with only wearable sensor data), while the perfect body region condition assumes that the active body region is always known accurately (i.e., if a procedure corresponds to the head, then the system correctly identifies the head as the active region). The estimated body region condition uses the approach described in Section III-B. The random forest and majority vote methods are analyzed within each body region condition.

Two hypotheses are evaluated using the clinical procedure detection system's results. Hypothesis $\mathbf{H_1}$ predicted that knowing the active body region will result in at least a 10% classification accuracy increase over not knowing the body region, while Hypothesis $\mathbf{H_2}$ predicted that the majority vote method will increase the random forest classification accuracy by at least 10%.

TABLE I

CLASSIFICATION ACCURACY (%) BY PROCEDURE, KNOWN BODY REGION CONDITION, AND CLASSIFICATION METHOD: RANDOM FOREST (RF) AND MAJORITY VOTE (MV).

Procedure	Body Region Condition					
	Unknown		Perfect		Estimated	
	RF	MV	RF	MV	RF	MV
IO Medication	0.00	0.00	0.05	0.00	0.00	0.00
IV Medication	0.12	0.27	0.37	0.36	0.03	0.00
Bagging	0.43	0.71	0.86	0.86	0.48	0.33
Blood-Pressure Cuff	0.03	0.00	0.39	0.60	0.12	0.50
CPR (Breath)	0.17	0.18	0.30	0.23	0.32	0.66
CPR (Compressions)	0.96	1.00	0.99	1.00	0.21	0.33
Chest-Tube	0.02	0.00	0.42	0.57	0.32	0.66
Combat Gauze	0.37	0.25	0.01	0.00	0.00	0.00
Combat Tourniquet	0.12	0.00	0.52	0.75	0.03	0.00
Draw Medication	0.20	0.20	0.47	0.47	0.32	0.66
ECG Leads	0.12	0.20	0.38	0.40	0.27	0.33
IM Administration	0.03	0.10	0.05	0.10	0.05	0.00
IO Line	0.14	0.29	0.61	0.86	0.15	0.00
IV Line	0.02	0.00	0.22	0.30	0.04	0.00
Intubation	0.27	0.33	0.49	1.0	0.28	0.66
King Airway	0.02	0.00	0.08	0.20	0.02	0.00
Oral Airway	0.09	0.08	0.27	0.33	0.00	0.00
Pulse-Ox Monitor	0.02	0.00	0.48	0.80	0.00	0.00
Splinting	0.13	0.00	0.80	1.00	0.18	0.33
Swab Area with Alcohol	0.00	0.00	0.12	0.13	0.06	0.00
Tie IV Tourniquet	0.03	0.00	0.17	0.11	0.01	0.00
Vital Monitoring	0.71	0.80	0.74	1.00	0.14	0.00
Wrap Head Wound	0.04	0.20	0.39	0.40	0.12	0.33
Average	0.18	0.19	0.40	0.50	0.14	0.21

ECG: Electrocardiogram and **IM**: Intramuscular

IV. RESULTS

The classification accuracy by procedure and known body region type are presented in Table I. Overall, CPR (Compressions) tended to be classified accurately the most, followed by bagging. These accurate classifications were due to the procedures' repetitiveness (i.e., chest compressions or squeezing the bag-valve mask). Vital monitoring was classified accurately as well, due to the procedure requiring minimal arm movements. Short-duration procedures, (i.e., oral airway or swabbing an area with alcohol), were difficult to classify and were often misclassified as a longer-duration procedure. Additional training data will potentially increase classification accuracy for short-duration procedures.

The classification accuracies corresponding to the unknown body region condition serve as a baseline condition, as no contextual data was used. The random forest method and majority vote method achieved an average classification accuracy of 18% and 19%, respectively. The majority vote method increased classification accuracy by at least 10% over the random forest method for five procedures, while two procedure's classification accuracies decreased.

Knowing the active body region with perfect precision increased classification accuracy dramatically for the random forest and majority vote methods, as the methods achieved an average classification accuracy of 40% and 50%, respectively. There was at least a 10% accuracy increase from the unknown body region condition for seventeen procedures using the random forest method and for nineteen procedures with the majority vote method. Both methods experienced a substantial decrease in accuracy for the combat gauze procedure. The majority vote method increased classification accuracy by at least 10% from the random forest method for nine procedures, while no procedure accuracy decreased by more than 10%. These results demonstrate that the majority vote method performs better than the random forest method, when the active body region is correctly identified.

Estimating the active body regions did not change the average classification accuracies dramatically from not knowing the active body region. Six procedures' random forest classification accuracies increased by at least 10%, while five procedures' accuracies decreased by at least 10%. The majority vote method using the estimated body region increased classification accuracy for ten procedures without knowing the body region, while seven procedures' accuracies decreased. Additionally, the majority vote method increased nine procedures' accuracies by at least 10% from the random forest method, while three procedures' accuracies decreased.

Overall, correctly identifying the active body region achieved the highest performance with both classification methods. Thus, illustrating the utility of using contextual information in activity recognition. The majority vote method achieved higher average classification accuracies than the random forest method, demonstrating the majority vote method's utility in a real-world complex environment.

V. DISCUSSION

Accurately detecting clinical procedures is critical, as a misclassification may result in incorrect patient care, and even death. The developed automatic clinical procedure recognition system did not produce accurate classifications. This result was expected due to the limited amount of training data and the unsophisticated approach to procedure detection. This preliminary work was meant to demonstrate how image data provides appropriate context that can improve a wearable sensor-based classification algorithm. Hypothesis $\mathbf{H_1}$ examined the impact of using image data to provide context to improve clinical procedure classification accuracy. The hypothesis is supported when the active body region is correctly identified without OpenPose. However, the hypothesis is not supported when the active body region is determined using OpenPose. The active body region detection method can be improved by incorporating multiple camera angles, as 3D representation of the medic's hands is feasible. Multiple camera angles may be less sensitive to object occlusion (i.e., the medic is blocking a camera view).

The developed body region detection method is also sensitive to the OpenPose skeleton keypoints, as the keypoints are a sparse representation of a human body. CPR (Compressions) were estimated frequently to be performed on the patient's head, when the compressions actually occurred on the chest. OpenPose has no chest keypoints, which generates the body region confusion. A machine learning algorithm may be trained using the two closest body parts for each hand in order to better estimate the active body region.

Assuming that a procedure's start and stop times are known may improve clinical procedure recognition, as a majority vote method may classify the procedure as a whole, instead of each individual window being classified. Hypothesis $\mathbf{H_2}$ tested the majority vote method's accuracy against the random forest accuracy, where each individual window is classified. The hypothesis is partially supported, as the majority vote method's unweighted average accuracy is greater than the random forest accuracy. However, the classification accuracy increased less than 10%. It is believed that the majority vote method will perform better in real-world scenarios, even without knowing a procedure's start and stop times. If seven out of twelve consecutive windows are classified as CPR, then the majority vote method will result in only CPR occurring in the twelve window time-frame. The random forest method will result in CPR and at least one other procedure occurring in the time-frame, which is most likely incorrect.

The planned future data collection will allow for a more sophisticated approach to clinical procedure detection. A larger training set will allow for deep learning algorithms to be applied, rather than the baseline signal processing methodology employed in this paper, where features can be learned from the wearable sensor data using convolutional neural networks. A long short-term memory recurrent architecture can be applied to the convolutional neural network to better capture the time-dependencies that occur within a procedure. Combining deep-learning techniques with the active body detection and majority vote methods is expected to improve the automatic clinical procedure detection system substantially. It is expected that future data collection will entail real-world environments in order to provide a more robust system validation.

VI. CONCLUSION

This paper used contextual information related to where the medic's hands are located relative to the patient, provided by image data, in order to improve clinical procedure detection accuracy. The developed clinical procedure detection system did not perform at the necessary medical domain standard, which was expected. The system is a necessary step towards achieving high performance, while demonstrating how contextual information and a majority vote method can be used in a complex real-world domain. Future work will improve the system's performance by incorporating deep learning and sophisticated image processing techniques.

REFERENCES

[1] M. K. James, L. A. Clarke, R. M. Simpson, A. J. Noto, J. R. Sclair, G. K. Doughlin, and S.-W. Lee, "Accuracy of pre-hospital trauma notification calls." *The American Journal of Emergency Medicine*, 2018.

[2] D. Tiel Groenestege-Kreb, O. Van Maarseveen, and L. Leenen, "Trauma team," *British Journal of Anaesthesia*, vol. 113, no. 2, pp. 258–265, 2014.

[3] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors." *IEEE Communications on Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.

[4] M. A. Panicker, H. Frigui, and A. W. Calhoun, "Cardio-pulmonary resuscitation (cpr) scene retrieval from medical simulation videos using local binary patterns over three orthogonal planes." in *IEEE Conference on Content-Based Multimedia Indexing*, 2018, pp. 1–6.

[5] J. N. Carlson, C. Karns, N. C. Mann, K. E. Jacobson, M. Dai, C. Colleran, and H. E. Wang, "Procedures performed by emergency medical services in the united states," *Prehospital Emergency Care*, vol. 20, no. 1, pp. 15–21, 2016.

[6] M. El Sayed, H. Tamim, and N. C. Mann, "Description of procedures performed on patients by emergency medical services during mass casualty incidents in the united states." *The American Journal of Emergency M edicine*, vol. 33, no. 8, pp. 1030–1036, 2015.

[7] F. Butler, *Tactical Combat Casualty Care Handbook (17-13)*. U.S. Army Center for Army Lessons Learned.

[8] S. Rawat, S. Vats, and P. Kumar, "Evaluating and exploring the myo armband," in *IEEE Conference on System Modeling & Advancement in Research Trends (SMART)*, 2016, pp. 115–120.

[9] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7291–7299.

Appendix M
Paper: Emergency Clinical Procedure Detection with Deep Learning

Emergency Clinical Procedure Detection With Deep Learning

Lingfeng Li[1], Richard A. Paris[1], Conner Pinson[1], Yan Wang[1], Joseph Coco[2], Jamison Heard[3], Julie A. Adams[4], *Senior Member, IEEE*, Daniel V. Fabbri[2], and Bobby Bodenheimer[1], *Senior Member, IEEE*

***Abstract*— Information about a patient's state is critical for hospitals to provide timely care and treatment. Prior work on improving the information flow from emergency medical services (EMS) to hospitals demonstrated the potential of using automated algorithms to detect clinical procedures. However, prior work has not made effective use of video sources that might be available during patient care. In this paper we explore the use convolutional neural networks (CNNs) on raw video data to determine how well video data alone can automatically identify clinical procedures. We apply multiple deep learning models to this problem, with significant variation in results. Our findings indicate performance improvements compared to prior work, but also indicate a need for more training data to reach clinically deployable levels of success.**

I. INTRODUCTION

Possessing and communicating accurate patient information is critical to achieving optimal medical outcomes. Unfortunately, too often emergency medical services (EMS) do not communicate complete information to a treating hospital [8]. This lack of communication can lead to inferior patient outcomes as the initial triage of a patient's condition can be done incorrectly at the receiving hospital [7]. This paper discusses a component of a noninvasive system that could potentially detect automatically what clinical procedures have been performed on a patient. Ideally, this system would supplement current care procedures by providing improved information on patient care to receiving hospitals, with the goal of improving a patient's initial triage level.

Heard et al. [7] made initial forays into this area by presenting a system that employed information from multiple types of sensors, including video, to categorize clinical procedures. Building on work in human activity recognition from multiple sensors [14], this work used contextual information provided by video sources to locate a medic's hands. Based on this location and other sensor information, this work represented a first step at clinical procedure identification. Its highest accuracy was achieved when the algorithm knew the active body region to which the procedure was applied (18% accuracy without body region knowledge and 40% accuracy with perfect body region knowledge).

This work was supported by the Department of Defense Contract Number W81XWH-17-C-0252 from the CDMRP Defense Medical Research and Development Program.

[1]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, Tennessee
[2]Vanderbilt University Medical Center, Nashville, Tennessee
[3]Department of Electrical and Microelectronic Engineering, Rochester Institute of Technology, Rochester, NY
[4]Collaborative Robotics and Intelligent Systems Institute, Oregon State University, Corvallis, Oregon
Corresponding author email: bobby.bodenheimer@vanderbilt.edu

This paper represents a new effort into this problem and only uses video data. It is motivated by recent developments in video classification and recognition in deep learning [5], [4], [9]. We use convolutional neural networks (CNNs) on raw video data to detect different clinical procedures performed during EMS transport. We attempted this because we want to see how far video data alone can take us, and to assess how much video data is necessary to achieve adequate performance from these data sets alone. There is an operational advantage from video data in that would not require paramedics to wear sensors. On the other hand, video data suffers from occlusion problems and noise due to lighting changes, and thus may have other difficulties for learning algorithms. Nonetheless, understanding how well video data by itself can work for classification is an important step in the development of an automatic clinical procedure system.

II. RELATED WORK

Human behavior recognition is the topic of a large amount of literature [13], and deep learning methods have been previously applied to the problem using video data [3] and non-video sensor-data [14]. When the problem is specialized to medical procedures, however, there is significantly less prior work [1], and this provides an opportunity for tuning pre-trained networks, as medical procedures have several significant identifying characteristics.

Karpathy et al. [10] demonstrated the effectiveness of convolutional neural networks (CNN) on several video classification tasks. They explored four different models for fusing information over temporal dimension through networks. All models exhibited strong capabilities for classifying video clips. Of their approaches, we primarily explore the single frame approach and the late fusion approach, as illustrated in Figure 1.

Ng et al. [9] and Donahue et al. [4] demonstrated the use of Long Short Term Memory (LSTM) for video classification tasks. These two groups of researchers processed individual frames with CNNs to aggregate information from frame data, and then the aggregated data is passed to the LSTM network for information summation. Results shown by Ng et al. and Donahue et al. suggest that LSTMs may achieve better video classification results than CNN methods if tuned well. Compared to the prior work, the limiting factor in the present work may be the relatively small amount of training data available.

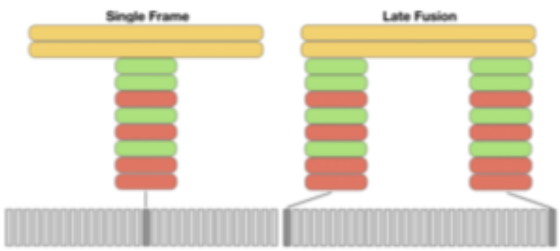


Fig. 1. Illustration of the Single Frame and Late Fusion Structures. The single frame model classifies based on one frame of the video stream (depicted at the bottom), whereas the late fusion model uses multiple temporally close frames. Red and green boxes together represent a mixture of convolutional layers and utility layers; yellow boxes represent fully connected layers.

Administer Medication	Bagging	Blood-Pressure Cuff
Chest-tube Prep	Chest-tube	Combat Tourniquet
CPR (Compression)	CPR (Breath)	Swab Area with Alcohol
Intubation	IO Line	IV Line
King Airway	Oral Airway	Pulse-OX
Draw Medication	ECG Leads	Vital Checking
Combat Gauze	Suturing	IM Administration
IV Tourniquet	Splinting	Wrap Head Wounds

TABLE I

Categories of clinical procedures. Abbreviations: IM – intramuscular; ECG – electrocardiogram; IV – intravenous therapy; IO – intraosseous; OX – oxygen.

III. Data Collection and Dataset Creation

All experimental data were collected in the Center for Experiential Learning and Assessment (CELA) at the Vanderbilt University Medical Center [11]. Seven subjects with medical training performed 24 different procedures (Table I) with varying amounts of repetition. Subjects performed the procedures on realistic medical mannequins that are commonly used for medical training. To collect video data four cameras were placed at different locations (Figure 2). They were positioned to capture as much of the procedure as possible and to ensure that the important parts of the procedure were always captured. Camera 2 (C2) was placed to emulate a ceiling mounted camera in an ambulance. Each camera collected video with a resolution of 3840×2160. This paper analyzed only C2 data as it was sufficient to ensure the mannequin, subject, and action were always visible.

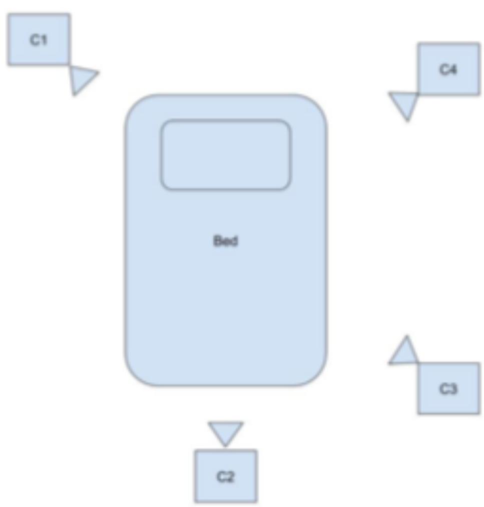


Fig. 2. Positioning of the four cameras used during data collection, as described in Paris et al. [11].

The collected video data were split into individual frames and each frame was assigned a category. Frames during which no procedure occurred were discarded. Each frame was cropped to reduce data size and eliminate extraneous information. Figure 3 shows the lines along which cropping occurred. The resulting frames were then resized to 256×256 pixels. In addition to the procedure name, each frame was labeled with the subject number and procedure occurrence as each procedure occurred multiple times.

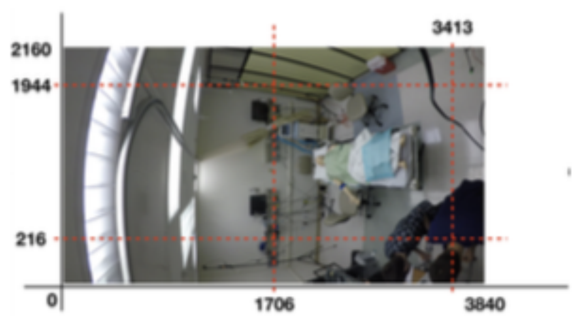


Fig. 3. Image cropping plan for frame data — only the central region is kept.

Of the data set composed of seven subjects, data from five subjects were used for training, one for validation, and one for testing. We performed 5-fold cross validation by rotating the subject used for validation and for testing. We chose to fold on subjects as we want to ensure the model generalizes to new and different medics performing the procedures.

Subjects in this experiment completed each procedure a randomized number of times and the time to complete each procedure varied leading to imbalanced classes. Class sizes ranged from 3,000 frames to 60,000 frames. To ensure the model does not overfit to one category and every class is equally represented the data was either downsampled or upsampled to approximately 8,000 frames. To downsample, approximately 8,000 random frames were selected and others discarded. To upsample, we duplicated each frame; then, if the upsampled category contained more than 8,000, it was downsampled as in the other categories. For example, in the fold 2 training set, the category *Pulse-OX* contained 3544 frames, and all frames were duplicated to reach 7088 frames (approximately 8000). Category *Chest-tube Prep* contained 6480 frames. We duplicated these frames and selected 8000 random frames, discarding the rest.

Validation and testing sets were balanced similarly with only the target number of frames changing. Our target was chosen so that only 15% of the classes would require duplication. The other 85% would then require downsampling. For each fold a different value was chosen based on the size of the categories in that validation or testing fold.

The previous balancing plan applies to two of the CNN models we test in this paper, the so-called "main model" (Section IV-A) and "variant 1" (Section IV-B). A third model, "variant 2" (Section IV-C), makes use of temporal relationships and has additional data inputs. For this model, we reduce our data size by resampling each session at 6 fps. In other words we choose frames 0, 5, 10, etc. To combat class balance issues we need to include more incidences of categories with lower representation. To do this we divide our categories into large, medium, and small. Large categories only resample once, medium categories resample an additional two times, and small categories an additional four times. These additional resamplings are offset from 1 to 4 frames from the first frame which gives us slight differences in our samples.

Our analysis system is built on Keras, Python's deep learning library.[1] Keras' built-in image augmentation framework helped increase the variability in the data. The brightness, rotation, zoom were all randomly modified. Images were randomly shifted vertically or horizontally, and could randomly be flipped vertically or horizontally. For models using temporal relationships, the same augmentations were applied to each related frame.

IV. Methods

The models in this section are largely based on InceptionV3 due to its success in image recognition tasks [12]. All training is done starting with the pretrained imageNet weights for low-level feature detection. For each fold the model was trained until a baseline validation accuracy was reached and the testing accuracy measured. This occurred three times per fold with the highest testing accuracy being recorded.

A. Main Model: Full Inception Model with Single Frame

Karpathy et al. [10] demonstrated that a single frame was sufficient to achieve a high accuracy (40%) in large datasets such as the UCF-101 and the Sport1M datasets. Those datasets are more complex than the clinical procedure dataset. The InveptionV3 architecture is a popular choice for deep learning image recognition tasks and we choose to use it as the basis of our network. Given the similarity of domains (action recognition), we expect similarly strong results. The clinical procedure dataset is smaller than similar datasets so the pretrained ImageNet weights were used to reduce training time and increase performance. Figure 5 is an illustration of this architecture, which we call our main model. Our model is made of the InveptionV3 architecture without the final output layer. Instead we place a fully connected layer, normalization, activation, dropout, and a second fully connected layer to adapt the model to our problem.

[1] https://keras.io/

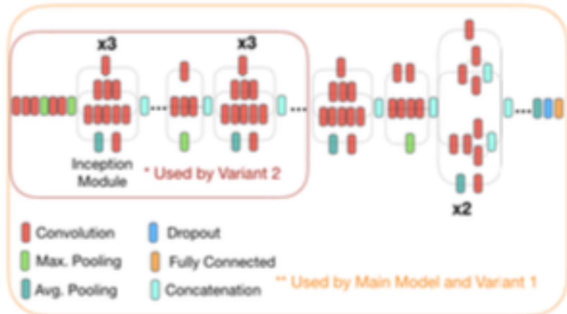


Fig. 4. A Compressed View of InceptionV3 [12]. The portion used by the main model and variant 2 are depicted with the orange and red lines, respectively.

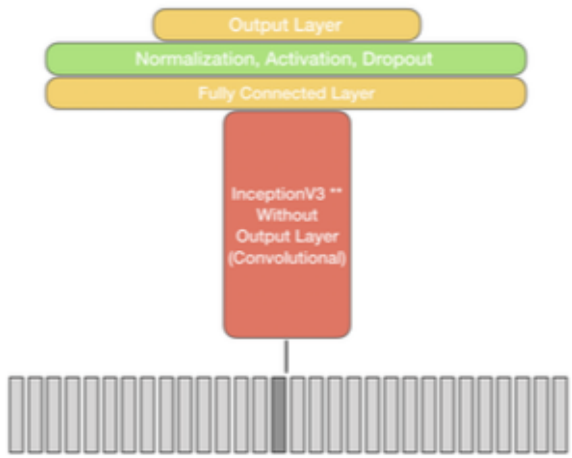


Fig. 5. An Illustration of main model, a direct adaptation of InceptionV3. A red box represents the part of model taken from InceptionV3 (portion of the model marked with ** in Figure 4); a yellow box represents dense layers; a green box represents utility layers.

B. Variant 1: Full Inception Model with Combined Categories

In the application domain for which we are training our recognition models, many of the procedures, while different actions, have temporal or physical associations with each other. For example, *Chest-tube Prep* must occur before *Chest-tube* and *CPR (Breath)* alternates with *CPR (Compression)*. In this variant we explore combining the *CPR (Breath)* and *CPR (Compression)* categories, as it is possible that we might get improved performance by grouping these associations manually rather than letting the algorithm determine them automatically. In particular, the main model commonly confused these two procedures and the goal is to remove that confusion and look for features common to both procedures.

C. Variant 2: Partial Inception Model with Late Fusion

Popular datasets, such as UCF101, Sport1M, or Youtube8M, cover a wide variety of topics, fields, and activities. Most of classes can be distinguished with one single frame. One problem we face is that some of the categories in the clinical procedure dataset, due to the similarity in context and equipment, cannot be easily distinguished from each other with a still image.

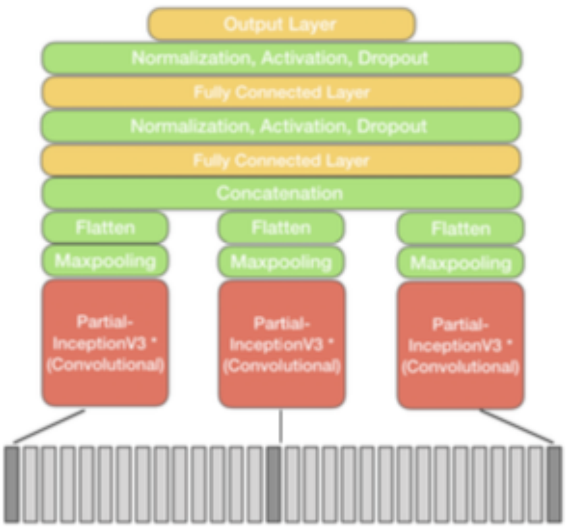


Fig. 6. The variant 2 model, a CNN model using the concept of Late Fusion. Red boxes represent the part of model taken from InceptionV3 (portion marked with * in Figure 4); yellow boxes represent dense layers; green boxes represent utility layers; light and dark grey boxes represent available and selected frame data.

To overcome this difficulty, we took inspiration from Late Fusion, and sample the video three times across a one second span: each data sample consists of the first frame of the second, the middle frame of the second, and the last frame of the second (1st frame, 15th frame, and 30th frame). Figure 6 is an illustration of our variant 2 (late fusion) model that takes advantage of both frame information and temporal information between frames. We first use convolutional layers and Inception Modules [12], which concatenate the results of different sizes of convolutional filters, to summarize information from each frame sample. Frame information is then fused to temporal relations in the top fully connected layers. Lastly, the output layer classifies clinical procedure based on given frame information and temporal information.

The standard InceptionV3 model is further pared down to use only six inception modules as a feature extractor. Three equivalent copies with shared weights of those six modules were created. These three were each fed into a separate maxpooling and flatten layer before being concatenated and fed through a series of layers as shown in Figure 6.

V. Results

Fold	Main Model	Variant 1	Variant 2
Fold 1	53.593%	57.798%	47.402%
Fold 2	42.398%	45.842%	50.506%
Fold 3	46.813%	44.975%	51.053%
Fold 4	44.525%	48.197%	61.770%
Fold 5	52.673%	53.504%	55.114%
Avg. Acc	48.000%	50.063%	53.169%

TABLE II

The testing accuracy of the three models across various folds and the resulting accuracy over all folds.

	Main Model	Variant 1	Variant 2
Administer Medication	36.743%	31.083%	12.222%
Bagging	66.476%	62.873%	90.371%
Blood-Pressure Cuff	39.218%	44.975%	24.572%
Chest-Tube	58.013%	59.649%	72.113%
Chest-Tube Prep	0%	0.422%	0%
Combat Gauze	45.668%	35.321%	35.576%
Combat Tourniquet	98.661%	94.474%	94.177%
CPR (Breath)	55.593%	72.120%	31.827%
CPR (Compression)	43.424%	N/A	32.649%
Draw Medication	7.099%	7.272%	4.382%
ECG Leads	56.050%	84.432%	69.910%
IM Administration	25.591%	18.931%	8.593%
Intubation	32.750%	51.789%	13.106%
IO Line	63.121%	67.852%	68.993%
IV Line	60.799%	58.810%	80.794%
IV Tourniquet	39.007%	35.527%	23.977%
King Airway	31.208%	47.505%	13.705%
Oral Airway	33.761%	41.290%	56.685%
Pulse-OX	35.269%	16.118%	10.000%
Splinting	79.053%	95.100%	61.038%
Suturing	54.862%	49.358%	40.258%
Swab Area W/ Alcohol	6.189%	19.863%	2.920%
Vital Checking	46.600%	24.816%	37.195%
Wrap Head Wound	92.089%	93.817%	81.202%
Total	47.474%	49.113%	41.731%

TABLE III

The testing accuracy by category from the 5-fold cross validation.

The results of the 5-fold cross validation can be found in Table II and the categorical accuracy of models across 5-fold cross validation can be found in Table III. Table III is computed by calculating the accuracy for each category for each fold and then averaging the categorical data across 5 folds, so that each experiment is given same weight in the result. Entry *CPR (Breath)* for variant 1 gives the accuracy for the combined CPR category. Table II is the traditional way of calculating accuracy for a cross-validation; however, Table III gives a better idea of the performance of the classifiers across each category, and since it balances each category equally, it gives a better overall indicator of performance.

The main model and variant 1 use single frame data for classification, while variant 2 uses 3 frames evenly spaced from 1 second of data for procedure detection. While they are using different types of data for classification task, they share the same goal of accurately detecting different clinical procedures. As a result, even though they use different types of testing datasets, their testing accuracies are generally comparable.

When comparing the overall results of Table II and Table III, the averaged categorical accuracy data of main model and variant 1 are within 1% difference with their testing accuracy data. This is because number of testing data for each category is approximately the same, and the 5-fold cross validation accuracy is a good estimator of the overall performance at making accurate classification for each model tested. However, there is a difference between variant 2's overall cross-validation performance and its averaged categorical accuracy data. This difference is caused by the low sample count of several categories. For example while the accuracy of category *Chest-tube Prep* and *Pulse-OX* is zero for Fold 4, category *Chest-tube Prep* only contains 9 samples and *Pulse-OX* only contains 19 samples, while other categories normally contains around 120 samples. Thus, this discrepancy supports the conclusion that our data set is too sparse to support high classification accuracy for some categories, and indicates which categories have sparse data sets.

VI. Discussion

The averaged categorical accuracy data of all three models are higher than the averaged accuracy achieved in the previous work where perfect knowledge of the body was assumed [7]. It therefore suggests that it is viable to perform the clinical procedure detection task without paramedics wearing sensors on their arms, although a combined method may yield higher performance that could reach standards high enough for the medical domain. Our results strongly suggest that richer video data would be helpful, and indicate where such data could be productively collected. In particular, we analyzed used data from only one video camera in this work. This was primarily done because the task of labelling and synchronizing the data among the cameras had only been completed for one camera when we began this work. However, it is likely that using the full video data record would lead to significant improvements and that is a course we are actively pursuing.

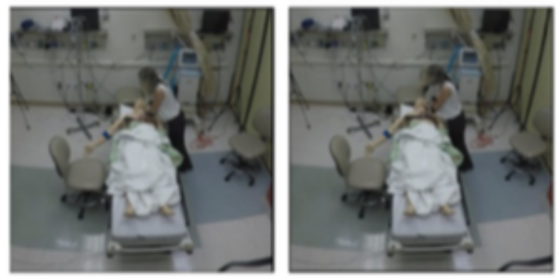
	CPR-B	CPR-C	Other	Accuracy
CPR-B Truth	**7064**	1825	3617	56.485%
CPR-C Truth	4084	**5029**	3416	40.139%

TABLE IV

Confusion Matrix Between CPR (Breath) and CPR (Compression)

Table IV is the confusion matrix between CPR (Breath) and CPR (Compression) for the main model. A confusion matrix summarizes the prediction of data according to their predicted labels. It is an accurate classification when the predicted label matches the true label (shown as bold figures in Table IV); otherwise, it is an inaccurate classification.

The table records all data from *CPR (Breath)* and *CPR (Compression)* categories (the first column) in the 5-fold and their predicted labels (the first row of the table). Predicted labels other than *CPR (Breath)* and *CPR (Compression)* are all recorded in *Other* category, and the accuracy for each category is calculated based on the given data. A close examination reveals that while each category achieves respectable categorical accuracy, a relatively large portion of the error is due to the model's inability to distinguish between *CPR (Breath)* and *CPR (Compression)*.



a. S1C2254560 CPR (Breath)   **b. S1C2254555 CPR (Compression)**

Fig. 7. Similarity between CPR (Breath) (left) and CPR (Compression) (right).

One cause of such inability is the close temporal proximity and the repetitiveness of the two categories. Compression and breath regularly happen one after another multiple times, and one major part of the CPR sequence is the transition time from one to the other. During data labeling, it is logical to randomly select a point in the transition and mark the division between two categories; however, during training, such an indistinct boundary between the two categories will cause confusion for the model. Figure 7 is an example of such problem. While both Figure 7a and Figure 7b come from the same transition period and they share significant similarity, Figure 7a is labeled as *CPR (Breath)*, but Figure 7b is labeled as *CPR (Compression)*.

Such confusion is avoidable. The hospital may not benefit from knowing how many times compression is applied versus that of breath, so it makes sense to combine the two categories and report CPR as one category, and the similar logic also applies to *Chest-tube* and *Chest-tube Prep*, discussed next.

Category *Chest-tube Prep* stands out among other categories, because it consistently has accuracy near zero for all models. Table V shows the sorted categories to which models wrongly classify data of *Chest-tube Prep*. The data suggests that for the majority of the time, models classify data from *Chest-tube Prep* as *Chest-tube* (59.678%). It suggests

Category	Percentage
Chest-tube	59.678%
IM Administration	24.552%
Suturing	10.866%
Others	4.904%

TABLE V

SORTED MISCLASSIFICATION CATEGORIES FOR CHEST-TUBE PREP.

that *Chest-tube* and *Chest-tube Prep* also suffer from the ambiguity problem during category transitions, as these two categories' occurrences are highly correlated. In addition, other factors such as the imbalance of data across subjects also potentially contribute to the low accuracy of *Chest-tube Prep*. There are 10,262 frames of data available for *Chest-tube Prep*; however, two out of seven subjects do not contain any data, while two subjects contains more than 6,000 frames of data.

As an aside, note that our classification procedure always categorizes an activity as belonging to one of the categories shown in Table I. In particular, there is no category for "no procedure" or anything analogous. This feature is by design, since it is unlikely that emergency medical personnel would be engaged in idle activity during transport of a patient to the hospital, and we felt it was unrealistic to include such a behavior in our data set.

From the model size's stand point, due to the fact that variant 2 (the late fusion model) only uses a partial InceptionV3 model up to "mixed6" layer, and the parameters are shared across all branches, the number of the parameters used by the this model ($\approx$ 14M parameters) is significantly less than that used by the main model ($\approx$ 24M parameters). The reduced number of parameters, however, does not suggest that variant 2 requires less computation power than main model.

Both variant 1 and variant 2 explore ideas to increase the performance of CNN based video classification methods for clinical procedures. The results and confusion matrices generated by main model, variant 1, and variant 2 give us suggestions for more practical ways of categorizing data and directions to create more powerful models.

VII. Future Work

In this work, we show that video classification methods from deep learning provide a tool for clinical procedure detection, potentially enhancing the communication between EMS and receiving hospitals. Future work includes training with improved data sets suggested by this work, incorporating data from multiple sensors into this framework as in Heard et al. [7], and trying improved deep learning models. The ultimate goal is to have models that work on video gathered in the field that can automatically detect clinical procedures to medically useful standards in real time, as discussed in Bloos et al. [2].

New deep learning models for video classification are continually emerging. As part of this work, we tried a CNN-LSTM model, similar to work by Ng et al. [9], but were not able to get satisfactory results. Also of note, Feichtenhofer et al. [6] have just made available their SlowFast network, which recently had high performance on a video benchmark. As deep learning models improve, we expect our own identification results to improve.

Acknowledgments

The authors thank Kaiwen Xu for helpful suggestions.

References

[1] M. S. Anju Panicker, H. Frigui, and A. W. Calhoun. Cardio-pulmonary resuscitation (cpr) scene retrieval from medical simulation videos using local binary patterns over three orthogonal planes. In *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–6, 2018.

[2] S. M. Bloos, C. D. McNaughton, J. R. Coco, L. L. Novak, J. A. Adams, R. E. Bodenheimer, J. M. Ehrenfeld, J. R. Heard, R. A. Paris, C. L. Simpson, et al. Feasibility assessment of a pre-hospital automated sensing clinical documentation system. In *AMIA Annual Symposium Proceedings*, volume 2019, pages 248–257. American Medical Informatics Association, 2019.

[3] J. K. Dhillon, Chandni, and A. K. S. Kushwaha. A recent survey for human activity recoginition based on deep learning approach. In *2017 Fourth International Conference on Image Information Processing (ICIIP)*, pages 1–6, 2017.

[4] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):677–691, 2017.

[5] S. Ebrahimi Kahou, V. Michalski, K. Konda, R. Memisevic, and C. Pal. Recurrent neural networks for emotion recognition in video. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, page 467–474, New York, NY, USA, 2015. Association for Computing Machinery.

[6] C. Feichtenhofer, H. Fan, J. Malik, and K. He. Slowfast networks for video recognition. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6201–6210, 2019.

[7] J. Heard, R. A. Paris, D. Scully, C. McNaughton, J. M. Ehrenfeld, J. Coco, D. Fabbri, B. Bodenheimer, and J. A. Adams. Automatic clinical procedure detection for emergency services. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 337–340, 2019.

[8] M. K. James, L. A. Clarke, R. M. Simpson, A. J. Noto, J. R. Sclair, G. K. Doughlin, and S.-W. Lee. Accuracy of pre-hospital trauma notification calls. *The American Journal of Emergency Medicine*, 37(4):620 – 626, 2019.

[9] Joe Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici. Beyond short snippets: Deep networks for video classification. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4694–4702, 2015.

[10] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[11] R. A. Paris, P. Sullivan, J. Heard, D. Scully, C. McNaughton, J. M. Ehrenfeld, J. A. Adams, J. Coco, D. Fabbri, and B. Bodenheimer. Heatmap generation for emergency medical procedure identification. In B. Fei and C. A. Linte, editors, *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10951, pages 779 – 785. International Society for Optics and Photonics, SPIE, 2019.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

[13] S. Vishwakarma and A. Agrawal. A survey on activity recognition and behavior understanding in video surveillance. *The Visual Computer*, 29(10):983–1009, 2013.

[14] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3 – 11, 2019. Deep Learning for Pattern Recognition.

APPENDIX N

Nashville Fire Press Release

Earlier this year, for 12-hour shifts over multiple months, two paramedics with the Nashville Fire Department worked with sensors placed on both wrists and both forearms, transmitting body motion and muscle activity to a server at Vanderbilt University Medical Center (VUMC). Meanwhile, a research-observer, seated inside the ambulance, logged the procedures performed on each patient. The VUMC, Vanderbilt University and Oregon State University research team is evaluating how well these data can be used to automatically produce clinical documentation.

Daniel Fabbri, PhD, Assistant Professor of Biomedical Informatics and Computer Science, is leading a project to improve the handoff of incoming emergency patients at military field hospitals and civil hospitals. The idea is to automatically generate patient acuity scores and abbreviated care records of in-transit patient procedures (e.g. CPR, intubation, etc.) based on computer interpretations of signals from sensors and video cameras.

On Nov. 18, at the American Medical Informatics Association Annual Symposium in Washington, D.C., Fabbri's team presented results of the feasibility work with Nashville Fire Department EMS.

"We wanted to start out with a small study to assess the technology and see what practical issues might arise, and this work with Nashville Fire proved quite fruitful. Due to privacy considerations we did not use video, but our initial findings bode well for the feasibility of our project," Fabbri said.

They call their system Automated Sensing Clinical Documentation because the system operates without medic input, using sensor data to produce documentation. As a result, medics can focus entirely on patient care without being distracted by writing down what they did. Simultaneously, the resulting documentation is extremely valuable for emergency room physicians and trauma surgeons who want to know what care has been provided.

"While our paramedics were outfitted with special armbands and watches, they didn't have to alter their daily routine or patient care protocols at all," noted Joaquin Toon, the EMS Quality Improvement Officer at Nashville Fire Department.

Fabbri's co-investigators for the feasibility study included colleagues from Biomedical Informatics (Dr. Laurie Novak), Emergency Medicine (Dr. Candace McNaughton), Anesthesiology (Dr. Jesse Ehrenfeld) and Electrical Engineering and Computer Science (Dr. Rorbert Bodenheimer), with partners at Oregon State University (Dr. Julie A. Adams). The project is supported by a $1.7 million research grant from the U.S. Department of Defense.

"Technology continues to advance. To think that a civilian paramedic or a military medic's hand and body movements can generate a patient medical record or alert the hospital of an incoming patient's condition is phenomenal. Nashville Fire Department was excited to partner with Vanderbilt Emergency Medicine in this research," said Commander Toon.

APPENDIX O

Quarterly Quad Charts

Automatic Sensing for Clinical Documentation (W81XWH-17-C-0252, DM160268)

PI: Daniel Fabbri **Org:** Vanderbilt University Medical Center **Award Amount:** $1,737,328.00

Study/Product Aim(s)

- Identify information needs at point-of-injury and trauma centers
- Develop a sensor system to generate EMR notes
- Pilot the system with civilian first responders

Approach

Develop a novel hands free clinical documentation system that leverages a combination of off-the shelf sensors, accelerometers, and cameras.

Images Courtesy Lt. Cmdr. J. Ehrenfeld — NATO Role III MMU

Timeline and Cost

Activities CY	18	18.5	19	19.5
Information Needs Gathering				
Sensor Development				
Sensor Prototyping				
Clinical Evaluation				
	169K			

Updated: 2018-01-14

CY18 Goals

- Information need observations and surveys
- Sensor development and testing in simulation lab

CY19 Goals

- Prototype development
- Clinical evaluation of data collection

Automatic Sensing for Clinical Documentation

DM160268
W81XWH-17-C-0252

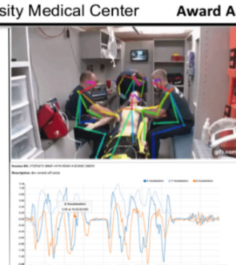
PI: Daniel Fabbri **Org:** Vanderbilt University Medical Center **Award Amount:** $1,737,328

Study/Product Aim(s)

- Identify information needs at point-of-injury and trauma centers
- Develop a sensor system to generate EMR notes
- Pilot the system with civilian first responders

Approach

Develop a novel hands free clinical documentation system that leverages a combination of off-the shelf sensors, accelerometers, and cameras. The output is a sequence of interventions performed.



Data collected and analyzed using video processing and accelerometer methods.

Timeline and Cost

Major Tasks CY	2017	2018	2019
1. Develop clinical detection algorithms using accelerometer data		39%	Draft & Submit Final Report
2. Develop clinical detection algorithms using image (video) data		27%	
3. Develop detection algorithms that combines accelerometer and image data		0%	
4. Design and implement high-level clinical activity features		0%	
5. Focus group & field data collection of developed prototype systems		26%	
Estimated Budget ($K)	$193k	$772k	$772k

Updated: 01 APR 18

Goals/Milestones

- CY18 Goals
  - ☐ Information need observations and surveys
  - ☐ Sensor development and testing in simulation lab
- CY19 Goal
  - ☐ Prototype development
  - ☐ Clinical evaluation of data collection

Comments/Challenges/Issues/Concerns

- Actual expenditure is lower than projected due to processing time of receipts from sub-contracts already underway, and initial spin-up time of project.

Budget Expenditure to Date

- Projected Expenditure: **$386,073**
- Actual Expenditure: **$178,186**

Automatic Sensing for Clinical Documentation

DM160268
W81XWH-17-C-0252

PI: Daniel Fabbri **Org:** Vanderbilt University Medical Center **Award Amount:** $1,737,328

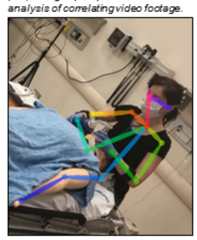
Study/Product Aim(s)

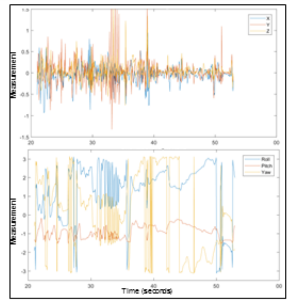
- Identify information needs at point-of-injury and trauma centers
- Develop a sensor system to generate EMR notes
- Pilot the system with civilian first responders

Approach

Develop a novel hands free clinical documentation system that leverages a combination of off-the shelf sensors, accelerometers, and cameras. The output is a sequence of interventions performed.

Graphs show accelerometer (top-right) and orientation (top-right) data collected from right hand while using a bag valve mask as seen in photo (left) using "open pose" software for analysis of correlating video footage.





Timeline and Cost

Major Tasks CY	2017	2018	2019
1. Develop clinical detection algorithms using accelerometer data		81%	
2. Develop clinical detection algorithms using image (video) data		75%	
3. Develop detection algorithms that combines accelerometer and image data		14%	
4. Design and implement high-level clinical activity features		17%	
5. Focus group & field data collection of developed prototype systems		56%	Draft & Submit Final Report
Estimated Budget ($K)	$193k	$772k	$772k

Updated: 01 JUL18

Goals/Milestones

- CY18 Goals
  - ☐ Information need observations and surveys
  - ☐ Sensor development and testing in simulation lab
- CY19 Goal
  - ☐ Prototype development
  - ☐ Clinical evaluation of data collection

Comments/Challenges/Issues/Concerns

- Expended 88% ($193,036) of projected funds ($218,224) during reporting period. Expenditures expected to continue to increase to programmed level as project expands data collection efforts.

Budget Expenditure to Date

- Projected Expenditure: $579,109
- Actual Expenditure: $396,410

Automatic Sensing for Clinical Documentation

DM160268
W81XWH-17-C-0252

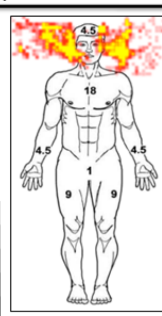
PI: Daniel Fabbri **Org:** Vanderbilt University Medical Center **Award Amount:** $1,737,328

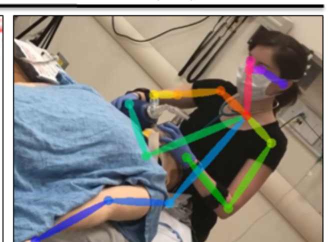
Study/Product Aim(s)

- Identify information needs at point-of-injury and trauma centers
- Develop a sensor system to generate EMR notes
- Pilot the system with civilian first responders

Approach

Develop a novel hands free clinical documentation system that leverages a combination of off-the shelf sensors, accelerometers, and cameras. The output is a sequence of interventions performed.





The illustration (left) is a heatmap depicting accelerator and video activity while performing an intubation while providing emergency trauma care in a simulation at Vanderbilt's Center for Experiential Learning and Assessment (CELA) lab (right).

Timeline and Cost

Major Tasks CY	2017	2018	2019
1. Develop clinical detection algorithms using accelerometer data		93%	
2. Develop clinical detection algorithms using image (video) data		85%	
3. Develop detection algorithms that combines accelerometer and image data		32%	
4. Design and implement high-level clinical activity features		27%	
5. Focus group & field data collection of developed prototype systems		56%	Draft & Submit Final Report
Estimated Budget ($K)	$193k	$772k	$772k

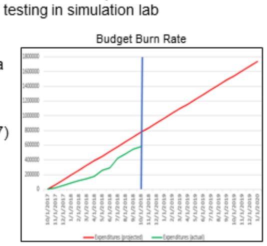
Updated: 01 OCT 18

Goals/Milestones

- CY18 Goals
  - ☐ Information need observations and surveys
  - ☐ Sensor development and testing in simulation lab
- CY19 Goal
  - ☐ Prototype development
  - ☐ Clinical evaluation of data

Comments/Issues

- Expended 94% ($182,259) of projected funds ($193,037) during this reporting period.
- Work progress on track.



Budget Expenditure

- Projected: $772,146
- Actual: $578,669

Automatic Sensing for Clinical Documentation

DM160268
W81XWH-17-C-0252

PI: Daniel Fabbri **Org:** Vanderbilt University Medical Center **Award Amount:** $1,737,328

Study/Product Aim(s)

- Identify information needs at point-of-injury and trauma centers
- Develop a sensor system to generate EMR notes
- Pilot the system with civilian first responders

Approach

Develop a novel hands free clinical documentation system that leverages a combination of off-the shelf sensors, accelerometers, and cameras. The output is a sequence of interventions performed.

(Above) A photo showing OpenPose software running an overlay while a doctor performs simulated CPR and then bagging in an ambulance at the Vanderbilt University Medical Center.

Timeline and Cost

Major Tasks CY	2017	2018	2019
1. Develop clinical detection algorithms using accelerometer data		99%	Draft & Submit Final Report
2. Develop clinical detection algorithms using image (video) data		98%	
3. Develop detection algorithms that combines accelerometer and image data		56%	
4. Design and implement high-level clinical activity features		47%	
5. Focus group & field data collection of developed prototype systems		63%	
Estimated Budget ($K)	$193k	$772k	$772k

Updated: 01 JAN 18

Goals/Milestones

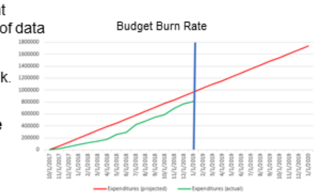
- CY18 Goals
  - ☐ Information need observations and surveys
  - ☐ Sensor development and testing in simulation lab
- CY19 Goal
  - ☐ Prototype development
  - ☐ System deployment
  - ☐ Clinical evaluation of data

Comments/Issues

- Work progress on track.

Budget Expenditure

- Projected: $965,182
- Actual: $808,664



Automatic Sensing for Clinical Documentation

DM160268
W81XWH-17-C-0252

PI: Daniel Fabbri **Org:** Vanderbilt University Medical Center **Award Amount:** $1,737,328

Study/Product Aim(s)

- Identify information needs at point-of-injury and trauma centers
- Develop a sensor system to generate EMR notes
- Pilot the system with civilian first responders

Approach

Develop a novel hands free clinical documentation system that leverages a combination of off-the shelf sensors, accelerometers, and cameras. The output is a sequence of interventions performed.

Apple Watch

Myo Gesture Control Armband

(Right) A photo showing a Emergency Medical Technician wearing the Myo and Apple Watch during system deployment with fire department while responding to patients to provide clinical care.

Timeline and Cost

Major Tasks CY	2017	2018	2019
1. Develop clinical detection algorithms using accelerometer data		99%	Draft & Submit Final Report
2. Develop clinical detection algorithms using image (video) data		98%	
3. Develop detection algorithms that combines accelerometer and image data		56%	
4. Design and implement high-level clinical activity features		47%	
5. Focus group & field data collection of developed prototype systems		63%	
Estimated Budget ($K)	$193k	$772k	$772k

Updated: 01 APR 19

Goals/Milestones

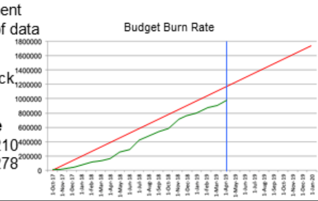
- CY18 Goals
  - ☑ Information need observations and surveys
  - ☑ Sensor development and testing in simulation lab
- CY19 Goal
  - ☑ System deployment
  - ☐ Prototype development
  - ☐ Clinical evaluation of data

Comments/Issues

- Work progress on track.

Budget Expenditure

- Projected: $1,158,210
- Actual: $960,278



Automatic Sensing for Clinical Documentation

DM160268
W81XWH-17-C-0252

PI: Daniel Fabbri **Org:** Vanderbilt University Medical Center **Award Amount:** $1,737,328

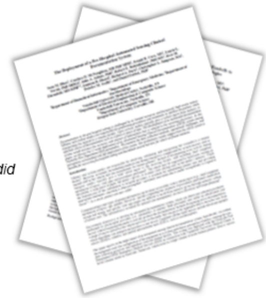
Study/Product Aim(s)

- Identify information needs at point-of-injury and trauma centers
- Develop a sensor system to generate EMR notes
- Pilot the system with civilian first responders

Approach

Develop a novel hands free clinical documentation system that leverages a combination of off-the shelf sensors, accelerometers, and cameras. The output is a sequence of interventions performed.

Project members submitted two articles this quarter to the American Medical Informatics Association (AMIA) 2019 Annual Symposium in D.S. from 16-20 NOV. AMIA accepted for oral presentation the paper, "The Deployment of a Pre-Hospital Automated Sensing Clinical Documentation System." AMIA did not accept the paper, "Understanding the Information Needs and Context of Trauma Handoffs to Design Automated Sensing Clinical Documentation Technologies" so the group will revise and resubmit.



Timeline and Cost

Major Tasks CY	2017	2018	2019
1. Develop clinical detection algorithms using accelerometer data		99%	Draft & Submit Final Report
2. Develop clinical detection algorithms using image (video) data		98%	
3. Develop detection algorithms that combines accelerometer and image data		60%	
4. Design and implement high-level clinical activity features			66%
5. Focus group & field data collection of developed prototype systems			88%
Estimated Budget ($K)	$193k	$772k	$772k

Updated: 01 JUL 19

Goals/Milestones

- CY18 Goals
  - ☑ Information need observations and surveys
  - ☑ Sensor development and testing in simulation lab
- CY19 Goal
  - ☑ System deployment
  - ☐ Prototype development
  - ☐ Clinical evaluation of data

Comments/Issues

- Requested No-Cost Extention

Budget Expenditure

- Projected: $1,351,255
- Actual: $1,215,201

Budget Burn Rate

Automatic Sensing for Clinical Documentation

DM160268
W81XWH-17-C-0252

PI: Daniel Fabbri **Org:** Vanderbilt University Medical Center **Award Amount:** $1,737,328

Study/Product Aim(s)

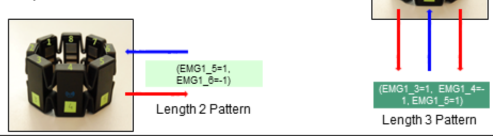
- Identify information needs at point-of-injury and trauma centers
- Develop a sensor system to generate EMR notes
- Pilot the system with civilian first responders

Approach

Develop a novel hands free clinical documentation system that leverages a combination of off-the shelf sensors, accelerometers, and cameras. The output is a sequence of interventions performed.

Working on new feature incorporating sequence data mining using the Myo to build a prediction model based on the set of patterns observed allow possible detection of clinical procedures without video.

- Hypothesis: Each procedure is characterized by a set of activity patterns.
- Process: Identify simultaneous activity patterns data. Single sensor pattern is positive or negative motion on one sensor. Multiple sensor pattern combines motions across sensors.



Timeline and Cost

Major Tasks CY	2017	2018	2019
1. Develop clinical detection algorithms using accelerometer data		99%	Draft & Submit Final Report
2. Develop clinical detection algorithms using image (video) data		98%	
3. Develop detection algorithms that combines accelerometer and image data			95%
4. Design and implement high-level clinical activity features			95%
5. Focus group & field data collection of developed prototype systems			95
Estimated Budget ($K)	$193k	$772k	$772k

Updated: 01 JUL 19

Goals/Milestones

- CY18 Goals
  - ☑ Information need observations and surveys
  - ☑ Sensor development and testing in simulation lab
- CY19 Goal
  - ☑ System deployment
  - ☐ Combine data for detection
  - ☐ Develop new features

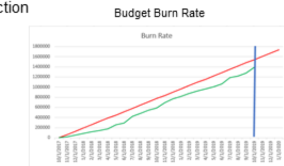
Comments/Issues

- Work progress on track.
- Requested No-Cost Extension

Budget Expenditure

- Projected: $1,544,292
- Actual: $1,392,415

Budget Burn Rate



Automatic Sensing for Clinical Documentation

DM160268
W81XWH-17-C-0252

PI: Daniel Fabbri **Org:** Vanderbilt University Medical Center **Award Amount:** $1,737,328

Study/Product Aim(s)

- Identify information needs at point-of-injury and trauma centers
- Develop a sensor system to generate EMR notes
- Pilot the system with civilian first responders

Approach

Develop a novel hands free clinical documentation system that leverages a combination of off-the shelf sensors, accelerometers, and cameras. The output is a sequence of interventions performed.

Sean Bloos was selected to orally present his paper "Feasibility Assessment of a Pre- Hospital Automated Sensing Clinical Documentation System." to the American Medical Informatics Association (AMIA) 2019 Annual Symposium in Washington DC on 18 NOV 19. This resulted in positive public engagement for the project both at the conference and in the press.

- https://mhealthintelligence.com/news/vanderbilt-researchers-test-mhealth-platform-to-improve-ed-hand-offs
- https://www.healthdatamanagement.com/news/documentation-system-seeks-to-improve-paramedic-ed-patient-handoffs
- https://www.ems1.com/technology/articles/nashville-fd-medical-center-testing-automated-documentation-system-for-er-handoffs-5lUXErraUVqGyl9/

Timeline and Cost

Major Tasks CY	2017	2018	2019	2020
1. Develop clinical detection algorithms using accelerometer data		99%		Draft & Submit Final Report
2. Develop clinical detection algorithms using image (video) data		98%		
3. Develop detection algorithms that combines accelerometer and image data				98%
4. Design and implement high-level clinical activity features			98%	
5. Focus group & field data collection of developed prototype systems			98%	
Estimated Budget ($K)	$193k	$772k	$772k	NCE

Updated: 29 July 20

Goals/Milestones

- CY18 Goals

☑ Information need observations and surveys
☑ Sensor development and testing in simulation lab

- CY19 Goals

☑ System deployment
☑ Develop new features

- CY20 Goals

☐ Evaluate combined data for detection

Comments/Issues

- Work progress on track.
- No-Cost Extension approved thru 29 DEC 20

Budget Expenditure

- Projected**: $1,470,047
- Actual**: $1,709,073

Budget Burn Rate

Automatic Sensing for Clinical Documentation

DM160268

W81XWH-17-C-0252

PI: Daniel Fabbri **Org:** Vanderbilt University Medical Center **Award Amount:** $1,737,328

Study/Product Aim(s)

- Identify information needs at point-of-injury and trauma centers
- Develop a sensor system to generate EMR notes
- Pilot the system with civilian first responders

Approach

Develop a novel hands free clinical documentation system that leverages a combination of off-the shelf sensors, accelerometers, and cameras. The output is a sequence of interventions performed.

Sean Bloos was selected to orally present his paper "Feasibility Assessment of a Pre- Hospital Automated Sensing Clinical Documentation System." to the American Medical Informatics Association (AMIA) 2019 Annual Symposium in Washington DC on 18 NOV 19. This resulted in positive public engagement for the project both at the conference and in the press.

- https://mhealthintelligence.com/news/vanderbilt-researchers-test-mhealth-platform-to-improve-ed-hand-offs
- https://www.healthdatamanagement.com/news/documentation-system-seeks-to-improve-paramedic-ed-patient-handoffs
- https://www.ems1.com/technology/articles/nashville-fd-medical-center-testing-automated-documentation-system-for-er-handoffs-5fUXEnnfJAqGy1k/

Timeline and Cost

Major Tasks CY	2017	2018	2019	2020
1. Develop clinical detection algorithms using accelerometer data		99%		Draft & Submit Final Report
2. Develop clinical detection algorithms using image (video) data		98%		
3. Develop detection algorithms that combines accelerometer and image data				98%
4. Design and implement high-level clinical activity features			98%	
5. Focus group & field data collection of developed prototype systems			98%	
Estimated Budget ($K)	$193k	$772k	$772k	NCE

Updated: 01 Apr 20

Goals/Milestones

- CY18 Goals

☑ Information need observations and surveys

☑ Sensor development and testing in simulation lab

- CY19 Goals

☑ System deployment

☑ Develop new features

- CY20 Goals

☐ Evaluate combined data for detection

Comments/Issues

- Work progress on track.
- No-Cost Extension approved thru 29 DEC 20

Budget Expenditure

- Projected**: $1,336,406
- Actual**: $1,673,353

Budget Burn Rate

Automatic Sensing for Clinical Documentation

DM160268
W81XWH-17-C-0252

PI: Daniel Fabbri **Org:** Vanderbilt University Medical Center **Award Amount:** $1,737,328

Study/Product Aim(s)

- Identify information needs at point-of-injury and trauma centers
- Develop a sensor system to generate EMR notes
- Pilot the system with civilian first responders

Approach

Develop a novel hands free clinical documentation system that leverages a combination of off-the shelf sensors, accelerometers, and cameras. The output is a sequence of interventions performed.

Sean Bloos was selected to orally present his paper "Feasibility Assessment of a Pre- Hospital Automated Sensing Clinical Documentation System." to the American Medical Informatics Association (AMIA) 2019 Annual Symposium in Washington DC on 18 NOV 19. This resulted in positive public engagement for the project both at the conference and in the press.

- https://mhealthintelligence.com/news/vanderbilt-researchers-test-mhealth-platform-to-improve-ed-hand-offs
- https://www.healthdatamanagement.com/news/documentation-system-seeks-to-improve-paramedic-ed-patient-handoffs
- https://www.ems1.com/technology/articles/nashville-fd-medical-center-testing-automated-documentation-system-for-er-handoffs-5tUXEnrafJAgGv1k/

Timeline and Cost

Major Tasks CY	2017	2018	2019	2020
1. Develop clinical detection algorithms using accelerometer data		99%		
2. Develop clinical detection algorithms using image (video) data		98%		
3. Develop detection algorithms that combines accelerometer and image data				96%
4. Design and implement high-level clinical activity features			98%	
5. Focus group & field data collection of developed prototype systems			98%	
Estimated Budget ($K)	$193k	$772k	$772k	NCE

Draft & Submit Final Report

Updated: 02 JAN 20

Goals/Milestones

- CY18 Goals

☑ Information need observations and surveys
☑ Sensor development and testing in simulation lab

- CY19 Goals

☑ System deployment
☑ Develop new features

- CY20 Goals

☐ Evaluate combined data for detection

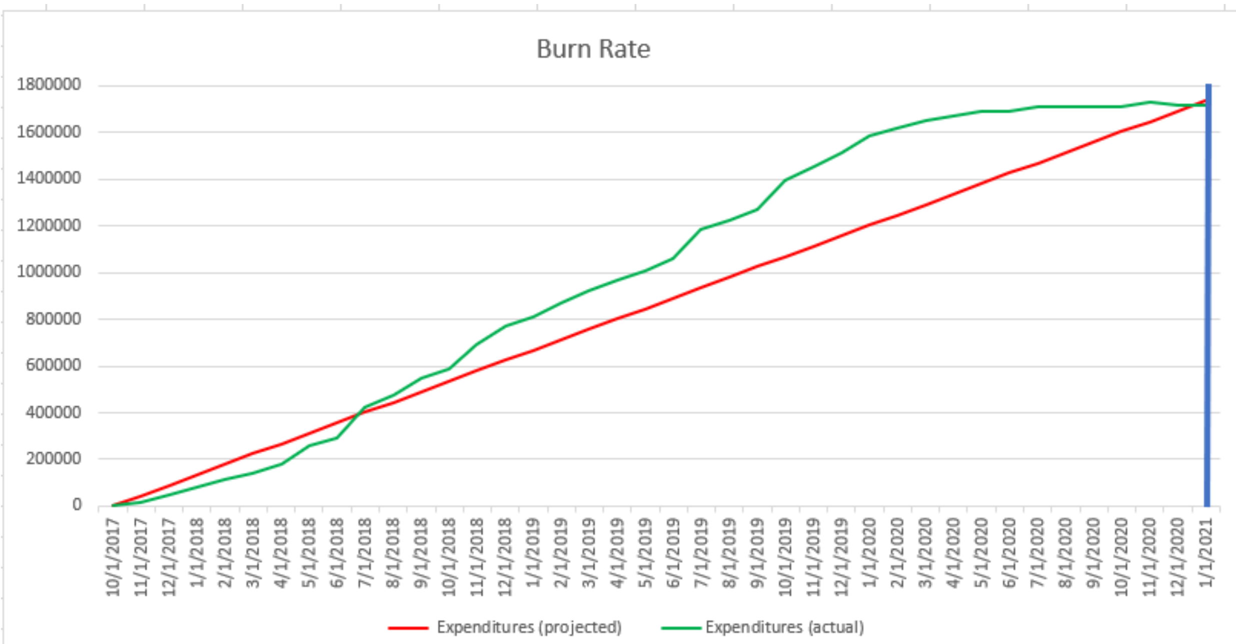
Comments/Issues

- Work progress on track.
- No-Cost Extension approved thru 29 DEC 20

Budget Expenditure

- Projected**: $1,336,406
- Actual**: $1,673,353

Budget Burn Rate



END OF REPORT