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 (e.g., 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 approved by the Institutional Review Board 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 (e.g., 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

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

Corresponding author email: jamison.r.heard@vanderbilt.edu

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

medical equipment was placed on the mannequin or on the 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 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.

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.

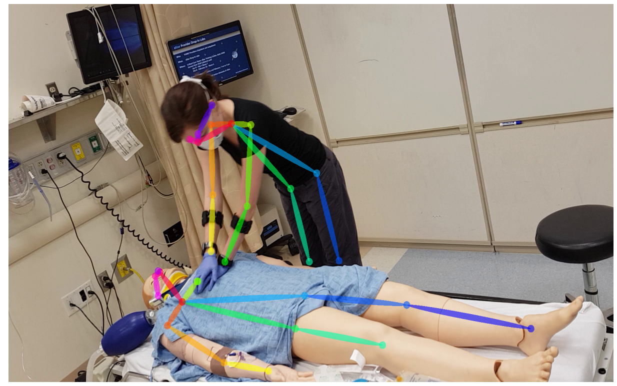

Fig. 1. OpenPose Output during CPR.

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 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

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))
```

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 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%.

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

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

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.