emsReACT: A Real-Time Interactive Cognitive Assistant for Cardiac Arrest Training in Emergency Medical Services

M Arif Rahman
Dept. of Computer Science
University of Virginia
Charlottesville, USA
mir6zw@virginia.edu

Leon Jia
School of Medicine
University of Virginia
Charlottesville, USA
lj4bd@hscmail.mcc.virginia.edu

Eimara Mirza
Dept. of Electrical and Computer Engineering
University of Virginia
Charlottesville, USA
em5nu@virginia.edu

Sarah Masud Preum
Dept. of Computer Science
Dartmouth College
Hanover, USA
Sarah.Masud.Preum@dartmouth.edu

Homa Alemzadeh
Dept. of Electrical and Computer Engineering
University of Virginia
Charlottesville, USA
ha4d@virginia.edu

Ronald D. Williams
Dept. of Electrical and Computer Engineering
University of Virginia
Charlottesville, USA
rdw@virginia.edu

John A. Stankovic
Dept. of Computer Science
University of Virginia
Charlottesville, USA
jas9f@virginia.edu

***Abstract*—EMS (emergency medical services) deals with cardiac arrest cases more frequently than any other fatal health conditions all over the world. We have developed emsReACT, a real-time interactive cognitive assistant, to train EMS providers for cardiac arrest cases in an emergency situation. This customized tool interacts in real-time with the first-responder and collects critical information. Using the conversational audio data available at EMS training sessions, emsReACT provides responder-specific decision support during the training based on domain specific information extraction, context-aware tracking of cardiac arrest protocols, and the dynamically changing condition of the patient. emsReACT leverages a dynamic behavioral model and a task-graph of two frequently used cardiac arrest EMS protocols.We have developed an intelligent abstraction mechanism with a critical risk-rating that drives an anytime algorithm to meet time requirements for regular and critical situations. Our thorough experimentation reveals an average end-to-end time of 2.7 seconds and 1.8 seconds for regular and critical interventions, thereby meeting the time requirements of 7 and 3 seconds, respectively. A qualitative study also reflects that over 70% of the 31 surveyed EMS providers rate the system as helpful to properly train the first-responders for executing cardiac arrest protocols.**

***Index Terms*—Emergency medical services, Interactive cognitive assistance, Intelligent Systems, Medical technologies.**

I. Introduction

Cardiac arrest is a complex, life-threatening health condition and one of the leading causes of death all over the world. In addition to the number of lives lost, cardiac arrest has a considerable economic impact as measured in terms of productive years of life lost due to premature death or other avoidable neurological disabilities [1]. Several factors can affect the outcome of an out-of-hospital cardiac arrest. One of which is the efficacy of emergency medical services (EMS) providers and first-responders who provide initial care to the suffering patient. To improve the quality of emergency healthcare in such crucial EMS scenarios, real-time interactive and assistive technologies should be adopted in EMS training sessions. However, case studies from the U.S. and Europe show that EMS training programs lack such automated cognitive assistants [2], and different phases of training are guided by manual interventions. Moreover, EMS scenarios vary in terms of degree of severity and complexity. A real-time cognitive assistant can contribute in multiple ways to improve the EMS training sessions for cardiac arrest protocols since the first responder would be physically working on a dummy and obtaining real-time feedback on their actions.

Key characteristics of cardiac arrest make the problem challenging. First, interventions relevant to the EMS cardiac arrest protocols are complex and must meet time constraints. Second, to follow the complex recovery procedure, first-responders need to recall critical information under a high-stress, overworked environment. This can lead to avoidable human errors [3]. Third, different interventions possess varying levels of severity, risk, and required degree of EMS training and expertise. Fourth, the importance of these factors also changes dynamically with time as the condition of the patient changes. For example, even some low-risk interventions might cause irreversible damage to patients if they are performed in an ill-timed or non-synchronized manner.

Addressing these characteristics of cardiac arrest lead to the following technical challenges:

- How to develop and implement a behavioral model of cardiac arrest protocols that match the dynamics of the patient recovery procedure. The model should demonstrate real-time situational awareness, i.e., it needs to reflect the dynamic information flow (e.g., the state of the patient) of an emergency cardiac arrest scene while interacting with first-responders within specific end-to-end time constraints. The dynamic information flow includes: (i) changing vitals, (ii) required medication dosage, (iii) varying degrees of risk, (iv) time-sensitivity and (v) dependencies between interventions.
- How to perform real-time and accurate concept extraction from conversational data on cardiac arrest which is unique for the EMS domain when compared to in-hospital medical and clinical text. This demands a specialized, domain-specific EMS lexicon to overcome the existing clinical concept extraction tools' limitations.
- How to perform real-time scheduling of a collection of collaborating tasks with dynamic deadlines driven by a risk factor. In addition, the solution should achieve acceptable performance under the effects of ambient noise at the scene, e.g., the noise of passing vehicles and bystanders' conversation.

Prior to creating a solution, we performed an empirical study conducted with EMS providers from local and regional EMS agencies. We found that automated and provider-customized feedback on the quality of physical interventions during EMS training should have significant positive impact on the skill development of the providers. For example, analyzing the training scene speech data from EMS providers to generate protocol specific feedback on interventions does not require any alterations during the incident, and creates lesser cognitive overload and better learning conditions for EMS providers.

To address the challenges and train first-responders properly for executing cardiac arrest protocols, we have developed **emsReACT** - A Real-Time Interactive Cognitive Assistant for Cardiac Arrest. Training in Emergency Medical Services. Note that since first-responders constantly communicate with each other during an scene, emsReACT is based on collecting and utilizing conversational data.

The main contributions of this emsReACT are:

- Developed and evaluated the first NLP based, real-time, and anytime cognitive assistant to provide automated, in-depth cognitive support in Emergency Medical Services (EMS) training sessions for time-sensitive and safety-critical cardiac arrest protocols. To the best of our knowledge, EMS still remains a novel domain for deploying and investigating an anytime automated assistant. Our research is the first one to address this scope.
- Designed a behavioral model and a task-graph as a state machine using the action-flow from the recovery procedure for two most frequently used cardiac arrest protocols. We deployed abstraction on the state-machine to solve the challenge of dynamic deadlines for generating feedback in different severity levels. We also introduced a risk-rating metric that dynamically controls an anytime algorithm to produce results in-time depending on the changing severity of the patient. Feedback in critical and regular situations have an average end-to-end response time of 1.8 s and 2.7 s respectively, both of which are within the requirements.
- For evaluation of emsReACT, we have collaborated with a regional EMS provider to get access to 12,000 textual narratives of real EMS scenarios. With direct participation of multiple EMS providers, we have recreated training exercises from 600 conversational textual cases. By injecting relevant types of noise profiles to mimic real EMS scenes in the audio data, we have evaluated different performance metrics of emsReACT.
- Experimented on noisy audio data to address the real-world issues and developed a resilient system that generalizes acceptably well under adverse situations. emsReACT outperforms benchmark tools such as MetaMap [4], cTAKES [5], and CLAMP [6] for the task of real-time information extraction specific to cardiac arrest cases. Considering the correctness, first-responders' expertise level, and timing, emsReACT feedback achieves an average F1-score of 87%.
- A survey of 31 EMS first-responders indicates that 23 of them mark the module as helpful for real-world cardiac arrest training. This provides strong evidence of the utility of the system.

II. Related Work and Background

A. Related Work

emsReACT addresses the problem of insufficient, real-time automation techniques in EMS training; and proposes an interactive, real-time, first-responder specific solution using training scene audio data. Authors in [7] leverage augmented reality and virtual reality based technologies for EMS training. However, we argue that emergency scenarios may have poor visibility issues and require real-time assistants, and the training phase should provide best surrounding conditions to the EMS providers. Using audio data eliminates visibility concerns. During the training, EMS providers go through various cognitive overloads in cardiac arrest related cases. Facilitating them with state-of-the-art real-time tools during training with minimum equipment overload can significantly improve the quality of the rescue task. Authors in [22] developed a method which presents pattern-based state-chart modeling approach for medical best practice guidelines such as model medical guidelines with basic state-chart elements. As this method is often not adequate for guaranteeing the correctness and safety of medical cyber-physical systems, and formal verification is required. To resolve the clinical validation aspect of the previous work, authors in [23] and [24] proposed an approach that transforms medical best practice guidelines to verifiable state-chart models and supports both clinical validation in collaboration with medical doctors and formal verification. However, none of these approaches adhere to the real-time dynamic aspect for any critical protocols. Previous studies [8] suggest that automated, real-time assistants for EMS training will ensure improvement of rescue quality. Authors in [9], [11] have addressed the challenges discussed in this paper, but they do not provide a real-time solution that scales for different level of expertise of the providers. Although there exist a few assisting systems for emergency response, most of them are generic and lack depth for any specific purpose. Sensitive cases such as cardiac arrest require extensive details and analysis in training sessions to prepare the EMS providers for real-world scenarios. emsReACT is a context aware real-time assistant that addresses this specific domain by assessing the clinical condition using training-scene audio data, and dynamically interacting with the EMS providers in real-time during EMS training. Following sections highlight related cognitive assistants from relevant domains, and how emsReACT is unique from existing literature.

B. Background on Cardiac Arrest

1) Cardiac Arrest Protocol: There are four different forms of cardiac arrest - ventricular fibrillation (VF), non-perfusing ventricular tachycardia (VT), asystole (A) and pulseless electrical activity (PEA) [12]. In this paper, we use the recovery protocols for two of these types of cardiac arrest - Ventricular Fibrillation (VF), and Pulseless Electrical Activity (PEA). The recovery protocols for these two types of cardiac arrest are complex and dynamic. A partial segment of two frequently used versions of the recovery process for the cardiac arrest protocol is depicted in Figure 2. For emsReACT, we use this standard EMS recovery protocol as the underlying model of a real-time feedback system. According to our EMT collaborators, each of the actions and interventions must be carried out in a timely manner for both of these protocols. The collaborators decided the time requirements to be a maximum time delay of 7 seconds for regular interventions and 3 seconds for critical interventions.

2) Intervention Risk and Certification Level of EMS Providers: EMS providers have different certifications, and they are allowed to perform different types of interventions. For example, there are two categories of cardiopulmonary resuscitation (CPR) training for healthcare providers and professional rescuers: (i) Basic Life Support (BLS), and (ii) Advanced Life Support (ALS) or Advanced Cardiac Life Support (ACLS). BLS providers are experienced with skills of scene safety, patient assessment, CPR by chest compressions, breathing, use of an automated external defibrillator (AED) and bag valve mask (BVM). EMT-basic providers are considered BLS. Compared to BLS providers, ALS or ACLS providers may give injections, administer medications, and place advanced intubation or airways - such as an endotracheal tube, laryngeal mask airway or esophageal-tracheal tube.

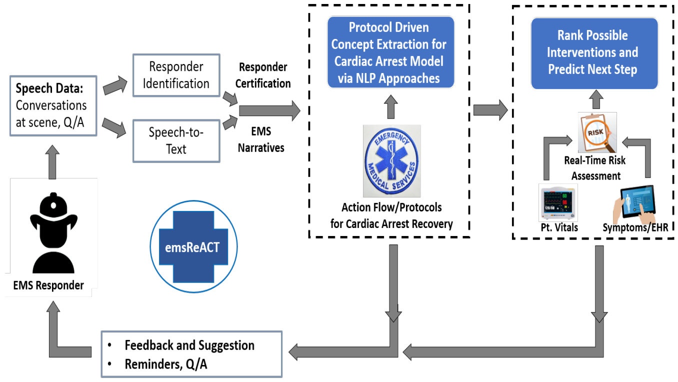

Fig. 1. emsReACT solution overview

EMT-advanced, EMT-enhanced and paramedics certification holders are ALS providers. Table I shows certification levels required for some of the interventions. For associated risks, a higher value indicates a higher risk. Risk-rating (O), risk-rating (NDWI), and risk-rating (DWNI) indicates associated original risk, risk if not done when indicated, and risk if done when not indicated, respectively.

TABLE I
SOME OF THE DYNAMIC RISKS AND REQUIRED CERTIFICATION LEVELS

Intervention	EMS certification level	Risk-rating (O)	Risk-Rating (NDWI)	Risk-Rating (DWNI)	Prerequisites/ Checks
12-lead ecg	Paramedic	1	4	1	BP, pulse, vitals
CPR	EMT-Basic	4	4	2	Allergies
intubation	EMT-Advanced	4	4	4	Allergies

III. Solution

emsReACT processes the training scene conversation of the care providers in real-time to understand the ongoing procedure, and provides suggestions and feedback. Specifically, the speech data is collected from the first-responder who is wearing a microphone. For each intervention, the first-responder is required to verbalize each of the actions for peer verification. Thus, using audio data from a training scene does not create any additional burden on the care providers. Figure 1 shows the high-level architecture of the system. The following subsections *A*,*B*, and *E* briefly describe the overall assistant and are included for completeness. The subsections *C* and *D* detail the main contributions of real-time dynamic scheduling for this paper.

A. Speech-to-text conversion

The first step of our solution is speech-to-text conversion in real-time. There is a lot of noise in EMS scenes, and the accuracy of transcriptions are significantly affected under such noisy conditions [10]. In the experiments of this paper, we consider both accurate and noisy transcriptions to reflect the potential variations in the performance of the off-the-shelf speech recognition tools. As this is not one of the main contributions of this paper, we do not detail the process here. We use the state-of-the-art Google Speech API for this step.

B. Concept extraction and context detection

Cardiac arrest related concepts are extracted in real-time from the speech, and converted to text as depicted in Figure 1. For extraction of concepts from the text, state-of-the-art clinical NLP tools, i.e. MetaMap [4], cTAKES [5], EMSContExt [14] and CLAMP [6] exist. However, these state-of-the-art tools are not best suited for real-time applications and they are not adapted for the EMS domain. In emsReACT, we use an EMS specific language model for detecting concepts from the speech using a lexicon expansion approach. We developed a detailed cardiac ontology [15], [16] to detect concepts from live speech data. A group of certified EMS providers helped us to develop a dictionary, D_1 with the following: (i) a specialized lexicon for cardiac arrest cases, (ii) a comprehensive vocabulary with a contextually mapped set of synonymous concepts and their possible homophones in noisy transcripts, and (iii) the related conditions/intervention prerequisites that might occur before/during the scene. We develop a bidirectional encoder representation from a transformer (BERT) based model for automated lexicon expansion and create another domain specific dictionary, D_2. Using binary classification on the dictionaries D_1 and D_2 in real-time, cardiac concepts are extracted from the speech narratives. We omit further details here as this is not our main contribution for the overall system.

C. Task abstraction for scheduling an anytime feedback

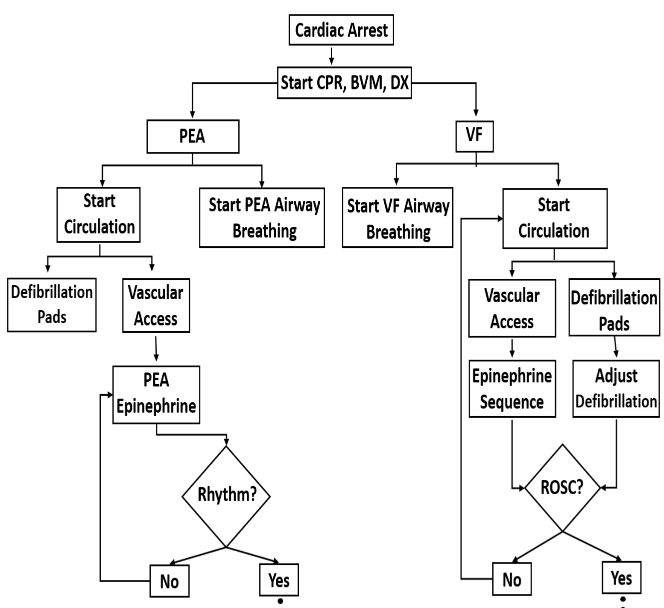

Fig. 2. Intervention flow (partial) for VF and PEA recovery

Cardiac arrest protocols do not follow any static flow of action, rather the overall procedure consists of many different dynamic actions or tasks (Figure 2). Implementing a system to adhere to the complexities of the interactions and associated time constraints is challenging. For example, most of the tasks are correlated with one another, however some of the tasks and dependencies are not mandatory. In addition, sometimes optional measures are also performed by the first responders for comprehensiveness of the patient recovery process. Critical tasks must always be carried out in a timely manner, while non-critical or optional tasks act as collaborative components for an improved patient recovery. The state of the patient which dynamically changes is the impetus for assigning a dynamic deadline to the collection of tasks. The dependency of the critical tasks must be carefully performed, but skipping the non-critical tasks and dependencies provide an option for the scheduling solution to adhere to the dynamically determined time-constraints. For emsReACT, we intelligently design the mandatory and optional nature of task correlation using an abstraction method [21]. This abstraction enables emsReACT to solve the dynamic time constraint issue and thereby providing real-time feedback to first responders for incidents with different severity.

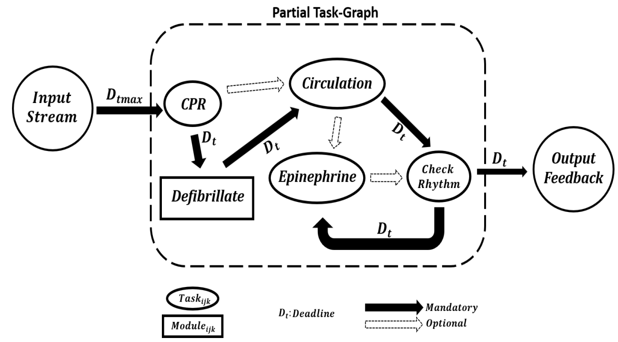

Fig. 3. Task abstraction concept for emsReACT

A key component of our solution is creating the task graph. The entire patient recovery process from Figure 2 must be converted into a task-graph with necessary abstractions for adhering to different time constraints, and how components depend upon each other, including different types of task collaborations. To provide some details, Figure 3 highlights the task-graph abstraction for a small portion of the recovery model. Here, the filled and dashed arrows indicate mandatory and optional task dependency, respectively. A task is denoted by an oval shape, and a set of related tasked is represented as a module in rectangular shape. For each $Task_{ijk}$ or $Module_{ijk}$, the associated properties i, j, and k denote whether the task/module is mandatory or optional (null task), the associated risk level according to current parameters or information, and the required list of information and pre-requisites, respectively. D_t denotes the dynamic deadline for the originating task. Depending on the severity and critical nature, this deadline updates dynamically for generating feedback through the *Output Feedback* step. We discuss a dynamic risk-rating based approach for updating the time-constraint deadline in the following subsection (subsection *D*). A potential feedback must be provided within this time-constraint for the associated task if any information or pre-requisite is missing in the input. If the time-constraint deadline permits, the optional route of the task-graph is explored for more comprehensive feedback. Otherwise, a prompt feedback is provided within the time limit using the limited available information. This type of scheduling method is uncommon in the literature in an application level, specifically when we have both "within" module anytime decisions and in-the-large anytime decisions "at the end-to-end" module level.

D. Real-time risk-rating assessment for situational awareness

To adhere to the different time-constraints for generating a feedback, we calculate a risk-rating via the anytime Algorithm 1. This rating indicates the current severity of the scene. The following criteria determine the dynamic risk-rating of the situation: (i) the set of allowed interventions by the acting EMS provider, (ii) the changing conditions of the patient, i.e., newly detected interventions and concepts and (iii) the dynamic risks associated with ongoing procedure. Table I shows the risks associated with each intervention, and how the severity of the situation changes when the care provider fails to carry them out in timely manner. Following the complex recovery procedure and dynamic task-graph illustrated in subsection *C*, and combining the current risk-rating with associated time constraint for each intervention, emsReACT calculates the sensitiveness of the situation in real-time. Then, the assistant provides feedback to the first-responders to meet the time requirements of 3 seconds for high-risk or critical conditions (risk-rating ¿ 7), and 7 seconds for low-risk or regular situations (risk-rating ¡ 7). If the deadline is 3 seconds, then emsReACT performs only the mandatory tasks and none of the optional, and when the deadline is 7 seconds the the system attempts to accommodate all of the tasks. The feedback component maximizes the accuracy of the automated response by allowing as much information as possible from the input audio stream within the time constraints. However, this timing constraint sometimes forces the algorithm to ignore some part of the remaining audio stream. Our experiments show that the critical cases sometimes lose additional information due to this time constraint. But for regular cases where the risk-rating is below 7, the anytime algorithm waits for the end of the intervention sub-task. The risk-rating and feedback deadline are constantly being monitored and updated with the change, update, or discovery of new scene concepts and interventions. For interaction between the real-time assistant and first-responder in the training, a list of frequently asked questions during EMS training for cardiac arrest cases is also provided to emsReACT. The first-responders can ask questions during the process and emsReACT can respond to those queries to minimize the cognitive overload of memorizing different steps.

Algorithm 1: Assessing situational awareness by dynamic risk-rating calculation via a form of an end-to-end Anytime Algorithm

Input: Streaming EMS audio, Concept list, Intervention flow for cardiac protocols
Output: Real-time feedback, risk-rating, severity of situation (*ratingRisk, situationVar*)

System Initialization
$conceptRisk \leftarrow 0$
$interventionRisk \leftarrow 0$
$ratingRisk \leftarrow conceptRisk + interventionRisk$
$situationVar \leftarrow 0$
while *Live audio stream is on* **do**
    if *New intervention is found OR Updated rating of previous intervention found* **then**
        Update interventionRisk
        Match with Intervention Flow
        Check rating Risk
        if $ratingRisk \geq 7$ **then**
            $situationVar \leftarrow critical$
            Output Urgent Feedback (within 3 s)
        else if *End of Intervention Sub-task* **then**
            Output Feedback (within 7 s)
        else
            if *Intervention is not carried out in time* **then**
                Update interventionRisk
                Check rating Risk
                Output Feedback (3s OR 7s)
            else if *Wrong intervention is carried out* **then**
                Update interventionRisk
                Check rating Risk
                Output Feedback (3s OR 7s)
    if *New concept found* **then**
        Update conceptRisk
    $ratingRisk \leftarrow conceptRisk + interventionRisk$
    if $ratingRisk \geq 7$ **then**
        $situationVar \leftarrow Critical$
        Output Urgent Feedback (within 3 s)
    else if $ratingRisk < 7ANDratingRisk > 0$ **then**
        $situationVar \leftarrow Regular$
        Output Feedback (within 7 s)

E. Personalized feedback generation for smart interaction

Different certification levels of care providers mandate the presence of multiple EMS providers in cardiac arrest related EMS training. When the acting EMS provider verbalizes intervention details for peer verification, emsReACT identifies the speaker and verifies the certification level of the EMS provider. This feature provides personalized feedback for specific level of EMS providers. Additionally, in some life-critical interventions such as CPR compressions, emsReACT uses speech identification technique along with the training scene transcriptions to provide a timely reminder for switching EMS provider to avoid exhaustion.

emsReACT is equipped with a speaker identification component which processes on scene conversation. To ensure the system is real-time, the model is trained with all the trainees before the beginning of the session. Different approaches exist in the literature for speaker identification. Sequence-to-sequence models are used for solving speaker identification problem, however the training phase is costly. Deep neural network based solutions are not effective for real-time EMS environment. We apply the basic method proposed in ARASID [19], this method is specially suitable for adverse conditions found during EMS training. Our experiments reveal following reasons for using this method: (i) ARASID identifies speakers using an artificial reverberation generator with different parameters to generate different artificial voice samples for each speaker. This means that it works well with limited training samples. (ii) The solution is easy to deploy, (iii) It filters out non-speech and overlapped speech samples, and separates non-trained speakers' samples. This feature means that the system filters out a large portion of background speech such as television speech, or an outside visitor. We do not detail the training method of ARASID for emsReACT due to space limitations.

IV. Evaluation Settings and Results

For evaluation or emsReACT, we have synthesized a dataset from real-world post incident EMS narratives obtained through our regional collaborator. EMS scenes were recreated for training exercises with multiple certified EMS providers in the laboratory settings. Techniques discussed in [11] were applied to develop synthesized dataset with noise-free audio, noisy audio, noise-free textual data, and noisy textual data. Although emsReACT takes audio streams as inputs, additional evaluation is conducted with the textual narratives to emphasize the robustness of emsReACT with respect to qualitative errors and different types of noises due to real-time transcription. Different styles of communication among the first-responders are also examined. We collected speech data from 14 EMT professionals along with their certification level while creating audio simulations to validate the accuracy of our speaker identification component. We used synonymous concepts, noise mappings and different homophones to enrich our specialized EMS lexicon [14]. Our dataset is created in a comprehensive fashion by considering audio, text and relevant noise profiles for training and testing different parts of emsReACT individually and in combination, i.e. accuracy and latency of speech to text conversation, cardiac concept detection, and quality of generated feedback in terms of generic and personalized nature. Time-sensitivity is added as a feature in deciding the accuracy. An ill-timed correct feedback is considered as false positive.

TABLE II
Description of synthesized dataset for emsReACT

Type	Description	Size/Samples
Text	EMS narratives	200
	Noise-inserted EMS narratives	200
Audio	Noisy audio (ambient noise)	20
	Noise-free audio	20
	Noise profiles	8
	Audio with artificial noise (using 8 noise profiles)	(20 X 8) = 160

A. Data Collection and Labelling

As live data collection in real EMS scenes requires certain approval and has privacy concerns, we collaborated with a regional EMS provider organization to collect the post-scene transcripts. We applied a style-transferring mechanism to recreate conversational data from these narratives. The annotations were supervised by certified EMS providers. Table II shows the sources, sizes, and types of our dataset. We have generated synthetic data by adding relevant audio and textual noise to original noise-free data [11]. However, some of our audio data originally had background noise. We have also used textual data from our regional collaborator. Each of the textual narrative samples comprises of 1000-1200 words, and the audio samples are 5-10 minutes long on average. To train emsReACT and different components of it, we have randomly selected half of each type of data shown in Table II, whereas the other half is used for test purposes.

B. Experimental Results

For the sensitivity of each intervention, correct timing of each feedback is an important element for emsReACT. The overall accuracy depends on the accuracy of each component. For example, if the speaker identification component did not detect the correct EMS provider and provided personalized feedback according to the wrong certification level, accuracy metrics record lower performance results. We also consider a correct, but ill-timed suggestion or reminder as false positive for evaluating the feedback system. Due to a lot of actual and simulated noise in our recreated EMS datasets, often parts of the original transcript gets distorted. This condition is the most contributing factor for overall lower accuracy numbers. Noise in audio sometimes leads to an indecisive state for emsReACT, different accents and communication styles adversely effect the speech recognition component. To demonstrate the applicability and time-sensitivity of emsReACT during EMS training sessions, here we show the accuracy of processing for concept detection, and generating an accurate feedback. We train the speaker identification component before the simulation, the transcription and speaker identification phase takes place concurrently. Table III shows the summary of overall accuracy. However, if the situation is detected as critical, emsReACT provides instantaneous feedback without further processing the transcription. This reduces the average time latency, but ignoring the remaining of the transcription causes the Precision, Recall and F-1 score to drop slightly. The minimum, average, and maximum end-to-end time for regular and critical feedback are 2.1, 2.7, 4.6 seconds, and 1.3, 1.8, 2.4 seconds, respectively.

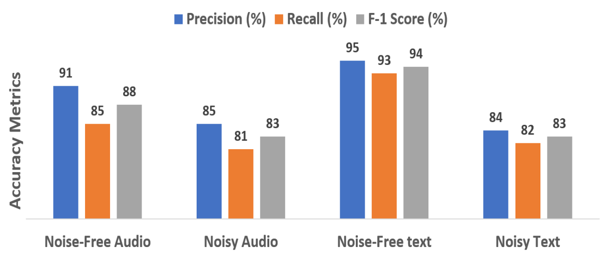

Fig. 4. Accuracy of emsReACT for different types of data

1) Comparison with existing methods for clinical concept detection and personalized feedback generation: State-of-the-art clinical information extraction tools such as MetaMap, cTAKES, and CLAMP work well with textual narratives. But these tools also process for other aspects of clinical contexts such as ranking, categorization and confidence scores. Thus the time required for detecting one specific concept is often too high for a real-time system. IMACS [9] provides feedback in real-time, however the feedback is generic for all the first-responders. emsReACT provides first-responder specific and customized solutions in real-time. Table IV shows the comparison of average F-1 score, and average time required for, (i) generating a feedback/reminder, and (ii) detecting a cardiac concept during EMS training, respectively from different types of data from our testing dataset for all state-of-the-art methods. emsReACT has the highest F-1 score of 91% (at least 8% higher compared to IMACS) and lowest average time of 2.7 seconds (at least 0.4 seconds lower compared to other approaches) to generate a generalized feedback in real-time and to detect a cardiac concepts, respectively. For generating a feedback personalized according to the expertise level of the current first-responder, emsReACT shows an F-1 score of 87%. emsReACT identifies the first-responder from speech, and uses a mapping that holds the certification level information for that specific first-responder for providing customized feedback. As IMACS does not generate personalized feedback, and MetaMap, cTAKES, CLAMP do not generate any feedback, we compare the accuracy of generalized feedback with IMACS and detection of cardiac arrest related concepts with all four of the methods.

Details of Precision and Recall scores are also listed in Table IV. emsReACT has at least 9% higher Precision and 3% higher Recall compared to the other approaches for detecting cardiac concepts. This is due to the generalization towards a wide range of noisy, real-world cases. emsReACT matches concepts from live narratives against a predefined vocabulary set listed with all possible cardiac arrest related concepts. This approach significantly reduces the false positives, and provides higher Precision scores. Compared to IMACS, we have also developed a mapping of homophones to the original cardiac concepts to ensure more resilience of emsReACT under noisy situations. The database we developed also consists of different pre-requisites of various interventions, and range of acceptable numerical quantities for intervention lengths and medication dosages for the cardiac symptoms. Using these information, emsReACT detects possible missing information and diagnosis while the training scene is ongoing, and provides crucial, decisive and timely feedback. This unique approach yields better Recall scores for emsReACT. For providing generalized feedback, emsReACT outperforms IMACS by at least 8% in Precision and by 7% in Recall. Training with a larger dataset increases the accuracy of our solution.

2) Performance of emsReACT for different types of data: To train our module, we have randomly selected half of each type of data shown in Table II. The other half is used for testing. The test dataset shows that for different types of data, average F-1 score is 87% (Figure 4) for generating the correct feedback specific to first-responder's expertise level. The error is mainly due to the inaccurate transcription from the speech-to-text engines, specially noisy surroundings affect emsReACT adversely. As we induce different noise profiles into the audios, the performance of emsReACT decreases. Some of the error is propagated due to out-of-flow actions by the first-responders. emsReACT detects only the interventions that are verbalized by them and recognized by the speech API. The inclusion of correct timing of feedback as a feature for determining accuracy metrics results in lower performance numbers. Low recall rate is contributed by some of the out-of-time feedback by emsReACT. Missing information from the conversational data creates a time-lag in the processing. emsReACT sends a wrong alert while waiting for the data, and consequently provides an incorrect feedback. Ill-timed correct suggestions are also resulted from such cases.

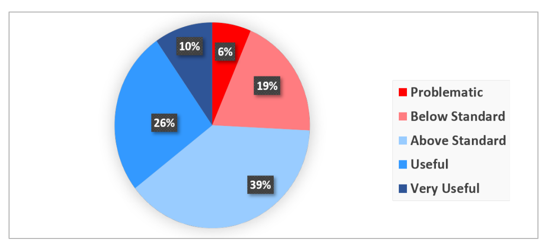

Fig. 5. Survey from 31 anonymous EMS providers

3) Qualitative Evaluation: emsReACT is also evaluated qualitatively by collecting anonymous EMS providers' responses using a Likert scale-based rating and open-ended interview. 31 EMS providers, who were not involved in the development phase, participated in the evaluation. For the overall idea and performance of emsReACT, 23 of the participating EMS providers consider the solution as either above standard, useful, or very useful as depicted in Figure 5. However, the remaining group of 8 EMS providers remarked that emsReACT might occasionally hinder care-providing when the provider interacts with it. Interestingly, in an open-ended interview, the later group also disagreed with the idea of using electronic devices and gadgets such as a microphone during EMS scene.

TABLE III
PERFORMANCE OF EMSREACT FOR PERSONALIZED ON-SCENE FEEDBACK AND TIME DELAY

Performance of emsReACT / Metrics	Average Latency of Each Sentence Level Subtask (s)		P	R	F-1
On-scene personalized feedback (regular)	Speech to text transcription via Google API	0.94 s	0.89	0.86	0.87
	Processing for concept and semantics detection	1.76 s			
On-scene personalized feedback (critical)	Speech to text transcription via Google API	0.57 s	0.78	0.71	0.74
	Processing for concept and semantics detection	1.24 s			

TABLE IV
COMPARISON FOR CARDIAC CONCEPT DETECTION AND/OR GENERATING FEEDBACK

Method	Avg. time (s)	Metric	Detecting cardiac concepts	Genera-lized feedback	Persona-lized feedback
emsReACT	2.7	P	95.14	93.72	88.89
		R	91.73	89.64	85.29
		F1	93.40	91.64	87.05
IMACS	3.11	P	85.91	85.01	N/A
		R	88.54	82.03	N/A
		F1	87.21	83.49	N/A
MetaMap	3.14	P	71.94	N/A	N/A
		R	63.21	N/A	N/A
		F1	67.29	N/A	N/A
CLAMP	3.21	P	65.21	N/A	N/A
		R	58.14	N/A	N/A
		F1	61.47	N/A	N/A
cTAKES	3.9	P	60.24	N/A	N/A
		R	63.95	N/A	N/A
		F1	62.04	N/A	N/A

The average year of EMS experience for the first and second groups are over 4 and 8 years, respectively. The demographic information collected in the beginning of the survey indicates that the second group of 8 EMS providers were less exposed to technological gadgets during their overall professional careers.

V. CONCLUSION

To the best of our knowledge, emsReACT is the first cognitive assistant that addresses the challenges of personalized, interactive decision support in EMS training. By utilizing an intelligent abstraction method in the recovery task-graph in real-time, emsReACT builds a collaborative pipeline of tasks that runs first without deadlines, and then dynamically identifies different timing constraints based on a novel risk factor. Importantly, this pipeline is not a static DAG (directed-acyclic-graph) and there needs to be a collaborative interaction between the elements of the pipeline. This combination of real-time challenges is not solved in the literature; thus, our solution is novel. Moreover, leveraging a novel data driven approach on the live speech data, emsReACT provides cognitive solutions for automated assessment of dynamic cardiac arrest related EMS training scenes. emsReACT provides customized feedback to the care providers according to their specific certification level in timely manner. Our thorough evaluation shows an average F1-score of 87% for personalized feedback generation in EMS training sessions in real-time, the average end-to-end time recorded for the feedback is 1.8 and 2.7 seconds for critical and regular cases respectively, which is within the acceptable delay span according to professional EMT personnel. emsReACT's accuracy is not 100% so it may sometimes provide wrong advice or feedback. However, it is not intended to work alone. Instructors work alongside emsReACT and can correct occasional errors. Extensive survey with 31 anonymous EMS providers reveal that emsReACT can play an important role in reducing the real-time cognitive overload. In the future we expect that emsReACT can also be used in actual EMS scenes. The methods discussed in this research can further be extended to address other complex system task-graphs, i.e., those found in systems that combine artificial intelligence and real-time solutions such as smart cities, smart health, etc.

1) Acknowledgement: This work is supported by the following awards: 60NANB17D162 and 70NANB21H029 from the U.S. Department of Commerce, National Institute of Standards and Technology (NIST).

REFERENCES

[1] Graham, Robert, Margaret A. McCoy, and Andrea M. Schultz. "Understanding the public health burden of cardiac arrest: the need for national surveillance." Strategies to Improve Cardiac Arrest Survival: A Time to Act. National Academies Press (US), 2015.
[2] Preum, Sarah Masud, Sirajum Munir, Meiyi Ma, Mohammad Samin Yasar, David J. Stone, Ronald Williams, Homa Alemzadeh, and John A. Stankovic. "A review of cognitive assistants for healthcare: Trends, prospects, and future directions." ACM Computing Surveys (CSUR) 53, no. 6 (2021): 1-37.
[3] Burnett, Susan J., Vashist Deelchand, Bryony Dean Franklin, Krishna Moorthy, and Charles Vincent. "Missing clinical information in NHS hospital outpatient clinics: prevalence, causes and effects on patient care." BMC health services research 11, no. 1 (2011): 1-7.
[4] Aronson, Alan R., and François-Michel Lang. "An overview of MetaMap: historical perspective and recent advances." Journal of the American Medical Informatics Association 17, no. 3 (2010): 229-236.
[5] Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & Chute, C. G. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association, 17(5), 507-513.
[6] Soysal, Ergin, Jingqi Wang, Min Jiang, Yonghui Wu, Serguei Pakhomov, Hongfang Liu, and Hua Xu. "CLAMP–a toolkit for efficiently building customized clinical natural language processing pipelines." Journal of the American Medical Informatics Association 25, no. 3 (2018): 331-336.
[7] Koutitas, G., Smith, S., Lawrence, G. (2021). Performance evaluation of AR/VR training technologies for EMS first responders. Virtual Reality, 25(1), 83-94.
[8] Daily, Bonnie F., James W. Bishop, and Robert Steiner. "The mediating role of EMS teamwork as it pertains to HR factors and perceived environmental performance." Journal of Applied Business Research (JABR) 23, no. 1 (2007).
[9] Rahman, M. Arif, Sarah Preum, John A. Stankovic, Leon Jia, Eimara Mirza, Ronald Williams, and Homa Alemzadeh. "IMACS-an interactive cognitive assistant module for cardiac arrest cases in emergency medical

service: demo abstract." In Proceedings of the 18th Conference on Embedded Networked Sensor Systems, pp. 621-622. 2020.
[10] Preum, Sarah Masud, et al. "Towards a cognitive assistant system for emergency response." 2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS). IEEE, 2018.
[11] Rahman, M. A., Preum, S. M., Williams, R., Alemzadeh, H., & Stankovic, J. A. (2020, April). GRACE: Generating summary reports automatically for cognitive assistance in emergency response. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 08, pp. 13356-13362).
[12] Parish, D. C., Goyal, H., & Dane, F. C. (2018). Mechanism of death: there's more to it than sudden cardiac arrest. Journal of thoracic disease, 10(5), 3081.
[13] Këpuska, Veton, and Gamal Bohouta. "Comparing speech recognition systems (Microsoft API, Google API and CMU Sphinx)." Int. J. Eng. Res. Appl 7, no. 03 (2017): 20-24.
[14] Preum, Sarah Masud, Sile Shu, Homa Alemzadeh, and John A. Stankovic. "Emscontext: EMS protocol-driven concept extraction for cognitive assistance in emergency response." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 08, pp. 13350-13355. 2020.
[15] Narayan, S. M., Wang, P. J., & Daubert, J. P. (2019). New concepts in sudden cardiac arrest to address an intractable epidemic: JACC state-of-the-art review. Journal of the American College of Cardiology, 73(1), 70-88.
[16] Muller, Dirk, Rahul Agrawal, and Hans-Richard Arntz. How sudden is sudden cardiac death. Circulation 114, no. 11 (2006): 1146-1150.
[17] Chapman, Wendy W., et al. "Evaluation of negation phrases in narrative clinical reports." Proceedings of the AMIA Symposium. American Medical Informatics Association, 2001.
[18] Finkel, J. R., Grenager, T., Manning, C. D. (2005, June). Incorporating non-local information into information extraction systems by gibbs sampling. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05) (pp. 363-370).
[19] Chen, Z., Ahmed, M. Y., Salekin, A., & Stankovic, J. A. (2019, March). ARASID: Artificial Reverberation-Adjusted Indoor Speaker Identification Dealing with Variable Distances. In EWSN (pp. 154-165).
[20] Hannun, Awni, et al. "Deep speech: Scaling up end-to-end speech recognition." arXiv preprint arXiv:1412.5567 (2014).
[21] Yao, Shuochao, Yifan Hao, Yiran Zhao, Huajie Shao, Dongxin Liu, Shengzhong Liu, Tianshi Wang, Jinyang Li, and Tarek Abdelzaher. "Scheduling real-time deep learning services as imprecise computations." In 2020 IEEE 26th international conference on embedded and real-time computing systems and applications (RTCSA), pp. 1-10. IEEE, 2020.
[22] C. Guo, Z. Fu, S. Ren, Y. Jiang, M. Rahmaniheris and L. Sha, "Pattern-Based Statechart Modeling Approach for Medical Best Practice Guidelines - A Case Study," 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), Thessaloniki, Greece, 2017, pp. 117-122, doi: 10.1109/CBMS.2017.14.
[23] C. Guo, S. Ren, Y. Jiang, P. -L. Wu, L. Sha and R. B. Berlin, "Transforming Medical Best Practice Guidelines to Executable and Verifiable Statechart Models," 2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS), Vienna, Austria, 2016, pp. 1-10, doi: 10.1109/ICCPS.2016.7479121.
[24] P. -L. Wu, D. Raguraman, L. Sha, R. B. Berlin and J. M. Goldman, "A Treatment Validation Protocol for Cyber-Physical-Human Medical Systems," 2014 40th EUROMICRO Conference on Software Engineering and Advanced Applications, Verona, Italy, 2014, pp. 183-190, doi: 10.1109/SEAA.2014.10.