TOWARD A MULTI-MODAL, INTERACTIVE AND SMART COGNITIVE ASSISTANT FOR EMERGENCY RESPONSE

M Arif Imtiazur Rahman

Charlottesville, Virginia

Master of Computer Science, University of Virginia, 2022

Bachelor of Science, Bangladesh University of Engineering and Technology, 2014

A Thesis submitted to the Graduate Faculty

of the University of Virginia in Candidacy for the Degree of

Doctor of Philosophy

Department of Computer Science

University of Virginia

August 2023

John A. Stankovic, Advisor

Lu Feng, Chair

Seongkook Heo

Homa Alemzadeh

Ronald D. Williams

Toward A Multi-modal, Interactive and Smart Cognitive Assistant For Emergency Response

M Arif Imtiazur Rahman

(ABSTRACT)

Emergency Medical Services (EMS) providers communicate extensively with many different stakeholders in emergency scenarios to ensure that the correct measures are taken and adverse outcomes are minimized. While communicating, the severity of the scene as well as the condition of the injured patients are often mentioned. Although state-of-the-art technologies such as noise-canceling microphones, smartwatches, and other devices aid the communication and recovery procedure, EMS training and providing care in emergency scenarios still remain very challenging and mostly manual-effort dependent. Most emergency scenes demand dynamic information flow, such as changing vitals, changing medication dosage, etc. which makes the task even more difficult. Previously, very few research have focused on building solutions that reduce the cognitive overload on the care providers, and provide interactive assistance based on the quality of the activity. This thesis presents novel research solutions for developing an automated cognitive assistant for EMS providers. Our research attempts to move the state-of-the-art toward a more comprehensive and automation orientated EMS intervention by utilizing natural language processing and transformer based language models on EMS textual corpus; and by effectively combining deep learning and attention mechanisms on data from smartwatch-based sensors and image data. The following research contributions with evaluations are presented. First, the thesis demonstrates the implementation of GRACE - a natural language processing based component to address formal documentation or reporting of critical information for emergency response. Second, the thesis presents an on-scene, data-driven, and protocol-specific framework, emsReACT, for interactive and personalized feedback to EMS providers during EMS training sessions and mock real-time incidents for cardiac arrest related cases. Third, a robust language model EMS-BERT is developed, for understanding the clinical concepts from live and existing EMS corpus. Fourth, two models SenseEMS and EgoCap are presented; for hand activity detection, monitoring, and real-time quality assessment, and a dataset development method for vision based EMS assistance, respectively. SenseEMS uses deep neural networks on smartwatch-based sensor data from the care providers. EgoCap dataset is developed by first-person captioning of images, which can be potentially used for scene understanding with contextual and visual features. The research results include working with regional EMS providers and certified EMS personnel, and involve real-life data collection and evaluation to show the effectiveness of each of the components. To summarize, the evaluation presented in this thesis successfully supports the hypothesis of the value of developing a cognitive assistant for EMS providers, and implies a successful feasibility of cognitive assistants for broader safety-critical domains.

Dedication

I dedicate my dissertation to Abu Nayeem, my family and many of my teachers and friends. My loving parents, M Habibur Rahman and Syeda Nure Kaseda - I left them for achieving this, I hope that they feel proud today. I sincerely thank my grandmother Karimunnesa and her brother A H Sharif and S H Sharif, my uncle M A Aziz, M Mujibur Rahman, Syed Rashel Kabir, Syed Saikh Imtiaz, and my sister Fariha Nur, who always believed in me. And, to my loving wife, Shabnam Wahed, who made countless sacrifices to make this journey of ours possible.

Acknowledgments

I would like to express my endless gratefulness to the Almighty Allah who has provided me with the opportunity to complete this work. I want to express my gratitude to all the people around me who encouraged me in various ways to complete this dissertation.

First, I would like to express my gratitude to my supervisor Dr. John A Stankovic, who is the best mentor I have ever been guided by, for guiding me throughout this journey. I am deeply moved and motivated by his continuous support, encouragement, and inspiration during the past five years. He has been the best adviser I could have ever imagined and I am grateful for his mentorship in all the projects I have endeavored on. I would also like to thank all the current and past members of Dr. Stankovic's research group who have directly or indirectly helped me with my research. I appreciate all their feedback, constant support, and help during data collection.

I would also like to thank the members of my dissertation defense committee: Dr. Lu Feng, Dr. Ronald D Williams, Dr. Seongkook Heo, and Dr. Homa Alemzadeh. I am very grateful to them and their research group for various help throughout this journey. I am also thankful to collaborators from University of Oxford and Aalborg University, particularly Dr. Zhuangzhuang Dai, Dr. Vu Tran, Dr. Andrew Markham, Dr. Niki Trigoni, and Dr. Chen Li for their support. The knowledge I gained by working with them is not only valuable for completing this work but also important for my future steps.

Last but not the least, I must thank and remember all the EMS responders who have been selflessly carrying out their duties and risking their lives to save people in distress. I particularly thank all the EMS professionals and responders including but not limited to Haydon Pitchford, Leon Jia, everyone from Richmond Ambulance Agency (RAA), Thomas Jefferson Emergency Medical Services (TJEMS), North Garden Fire Company (NGFC), and Virginia Office of the Emergency Medical Services (VaEMS), who were directly or indirectly involved in developing, testing, and surveying phases of this project. I also acknowledge the support from the U.S. Department of Commerce, National Institute of Standards and Technology (NIST) for the following awards: 60NANB17D162 and 70NANB21H029.

Contents

List of Figures

List of Tables

List of Abbreviations

ACLS Advanced Cardiac Life Support

AED Automated External Defibrillator

ALS Advanced Life Support

BERT Bidirectional Encoder Representations from Transformers

BLS Basic Life Support

BVM Bag Valve Mask

CC Chief Complaint

CPR Cardiopulmonary Resuscitation

DL Deep Learning

EMS Emergency Medical Services

HCI History of Present Illness

ML Machine Learning

NER Named Entity Recognition

NLP Natural Language Processing

NLTK Natural Language Toolkit

PEA Pulseless Electrical Activity

PMH	Past Medical History
VF	Ventricular Fibrillation

Chapter 1

Introduction

1.1 Motivation and Challenges

During emergency situations, first-responders collect, aggregate, filter, and interpret information from different sources. Processing such a huge amount of real-time information at the incident scene requires a significant amount of cognitive overload. This thesis presents key components of a cognitive assistant system for emergency care providers to reduce their on-scene cognitive overload by real-time collection and analysis of data; and by providing dynamic, data-driven decision support in an interactive manner. The components of the cognitive assistant leverage responder-worn devices, i.e., microphones, camera and wristwatch-based smart sensors to monitor activities and verbal communications, and aggregate the dynamic information with static sources, i.e., emergency response protocol guidelines to generate insights that can assist the providers with effective on-scene decision making and post-scene activity validation during training sessions. The high-level technical challenges addressed in this thesis are outlined below.

1. Semantic inference from EMS textual data, i.e., negation detection, temporal expression detection, and value association for accurate information extraction, and deep inference of EMS text, i.e., distinguishing patient-related information

from scene and chronological ordering of information for automated documentation.

2. Developing protocol-specific, data-driven behavioral models for interactive decision support during the EMS training. Context sensing and developing situational awareness using speech data from the scene for an automated EMS training assistant to provide customized and accurate decision support according to varying degree of responder expertise. The severity of each intervention and associated risk for not following a proper sequence of actions in timely manner must be considered to avoid adverse outcome;

3. Domain-specific entity extraction that is unique for emergency response when compared to current medical and clinical oriented ontologies, as the specialized vocabulary used by first-responders limits the applicability of existing solutions;

4. Fusion of sensor data to detect and verify the accuracy of gestures for hand operated interventions performed by the EMS providers. This is critical to detect deviations from standard metrics during training sessions for time-sensitive, safety-critical interventions, i.e., CPR;

5. Lack of an image dataset for modeling a first person or egocentric captioning of EMS scene to provide a precise perspective, as a third person narrative often imposes ambiguity. Egocentric vision data are often contaminated by noise caused by motion, occlusion, and awkward camera angles which affect the accuracy of extracting object attributes or key features.

1.2 Contributions

The first component of the assistant is developed with a vision to relieve the care providers from the cognitive burden of memorizing crucial facts related to the scene. This is achieved by automatically filling out the intervention summary report for the emergency first-responders. A tedious, cognitively overloaded task the care providers have to perform in any emergency scene is post operative manual report generation. These reports often lack critical details that are collected from the scene in real-time. Research shows that in the USA, 13.6% of the time mistakes are made while inputting information into the summary forms (Burnett et al. 2011). Mistakes happen in the form of inputting wrong information, forgetting to include a correct piece of information, and misplacing data in the wrong field of a form. Discussions with first-responders indicate that automatic form filling followed by only needing to check the forms would be a tremendous aid in their jobs.

The second contribution of this thesis is the development of an interactive, data-driven, and real-time framework for providing personalized feedback to the providers during EMS training sessions for cardiac arrest related cases. Previous studies suggest that automated assistants for EMS ensure improvement of intervention quality (Nicholson et al. 2017; Daily, Bishop, Steiner, et al. 2007). However, case studies in the USA show that EMS training programs lack such automated cognitive assistants (Pozner et al. 2004), and different phases of training are guided by manual efforts only. EMS providers perform time-sensitive, safety-critical interventions to ensure patient safety and reduce risk. Usually multiple EMS providers with different level of expertise and training certification engage in providing care to a single patient. Improving the quality of EMS interventions through interactive, assistive technology can contribute to improving the quality of overall emergency healthcare. Moreover, EMS scenarios vary in terms of degree of severity and complexity. Thus some scenarios require EMS providers to go through more rigorous training to obtain an adequate level of expertise. Utilizing audio data obtained from body-worn devices, i.e., microphones, this thesis presents a cognitive assistant which is situation-aware, and provides real-time, protocol-specific, personalized feedback to the provider for improved decision making during EMS training and selective mock incidents for cardiac arrest cases. Based on an empirical pre-study conducted with EMS providers from local and regional EMS agencies, the findings indicate that automated and provider-customized feedback with alert messages on quality of physical interventions during EMS training sessions can have significant positive impact on the providers' skill level.

The third part of the overall work focuses on addressing the challenges related to domain-specific entity and concept extraction. For extraction of clinical concepts, state-of-the-art clinical NLP tools, i.e. MetaMap (Aronson and Lang 2010), cTAKES (Savova et al. 2010), and CLAMP (Soysal et al. 2017) exist. These tools are designed for general purpose clinical concept detection from structured medical data. Preliminary experiments for this thesis reveal that these tools obtain poor performance for real-time EMS transcripts and EMS corpus. Existing bidirectional encoder representation from transformer (BERT) based clinical models such as BioBERT (Lee et al. 2020) is trained with structured textual corpus. EMS live transcripts require a concept extraction tool that scales well with unstructured dataset. For EMS specific purposes, this research develops a BERT based model, EMS-BERT, for automated detection of EMS domain related entity and concepts with lexicon expansion. Using domain specific dictionary developed from lexicon expansion, EMS-BERT performs concept extraction from the speech data collected at the scene. EMS-BERT also provides solutions for detecting missing data and relation extraction using intervention dependency static information.

The fourth part of the research, SenseEMS and EgoCap address multimodal assistance for EMS providers using smartwatch-based sensor data and first-person captioning of potential EMS scenes, respectively. Using accelerometer, gyroscope, and magnetometer data from the smartwatch, SenseEMS presents an ongoing research development for hand activity detection and monitoring, and verifying sensitive parameters of EMS interventions for quality improvement. This research tracks different hand gestures while performing critical hand operated interventions. The SenseEMS model generates both real-time and post-scene validation and quality assurance feedback for EMS providers to improve their skills. Currently, a human supervisor is appointed to verify all the ongoing hand activity during an EMS incident, and provide separate reports to acting responders about the quality of their performance. Previous research have discussed the drawbacks of manual supervision, and proposed using different sensors for assistance and task verification (Chang, Kang, and P.-C. Huang 2013; Rizzo et al. 2015; Sonntag 2015; Gamberini et al. 2009). However, none of them provide solutions for an automated system for evaluating the quality of any EMS hand operated interventions. Ongoing work on SenseEMS aims at creating an automated hand activity monitoring system to train EMS providers, and providing real-time verification of intervention quality. Results discussed in this thesis present an improved model for gesture detection during two hand operated EMS interventions, i.e. chest compressions, and attachment of defibrillation pads; and a real-time approach for detecting the rate of compression for one of the most the critical, life-saving EMS intervention - cardiopulmonary resuscitation (CPR).

The last part of the thesis develops an image dataset with first-person captioning of daily scenes which are relatable to EMS environments. This dataset, EgoCap, can be highly influential for automated understanding of the scene and it's relevant objects. First-person captioning is significant for EMS as it provides a machine vision of the notion of "self". It is important to depict the scene from an egocentric perspective with an emphasis on the responder's status, activity, and position. Ego-captioning is non-trivial since Ego-images can be noisy due to motion and angles. Besides, describing a scene in a first-person narrative involves drastically different semantics. Different empirical implications have to be made on top of visual appearance as the cameraperson is often outside the field of view. EgoCap notes that ego-captions are often accredited to contextual cues, such as when and where the event unfolds, and whom the first-responder is interacting with. This inspires the fusion of contexts for situation-aware captioning as well for an EMS scene. Different aspects of creating the EgoCap dataset and it's properties are described in this section of the thesis.

To summarize, the thesis discusses the following solutions for creating a cognitive assistant for EMS training and mock-real scenes, and creating a common EMS platform by integrating the following tools: *(i)* formal documentation of patient summary reports from EMS scenes, *(ii)* an automated assistant for customized interactive feedback and suggestions for cardiac arrest related training, *(iii)* a language model for EMS concept/entity detection from EMS corpus, *(iv)* sensor data fusion for providers' hand activity detection and monitoring, and ego-captioning of vision data with enhanced contextual awareness for EMS. The primary list of contributions and novelties of this thesis are outlines as following.

- Developed the first natural language processing based system to address formal documentation or reporting of critical information for patients in an EMS scene. The assistant addresses the semantic challenges (i.e. information synchronization, value association, negation detection) in EMS text and information validation (i.e. vitals) for EMS data under both noise-free and noisy conditions.

- Developed a cognitive assistant for providing real-time interactive feedback to multiple EMS providers during EMS training and mock-real scenes using the conversational audio data. The thesis presents a protocol and data-driven, responder-specific approach to provide decision support during the scene, and create a standard platform for EMS training.

- Developed a transformer based domain-specific language model for EMS domain that extracts EMS concepts/entities and detects missing information from EMS corpora. The model combines ontology based lexicon expansion approach with semantic heuristics and inferences for entity detection and relation extraction.

- Developed a hybrid deep neural based solution to analyze gesture detection for EMS hand operated interventions, i.e., CPR and defibrillation, using smartwatch-based sensor data from the providers. The ongoing work provides real-time metric updates and detects deviations from standard procedures.

- Developed a unique dataset through egocentric captioning of EMS-relevant daily-life based images from different public image datasets. The new dataset fuses the contextual knowledge using first-person captioning to provide better understating of the scene for the EMS providers. In the future, the dataset will be used to train a transformer-based network with visual-context fusion modules to automate ego-captioning of EMS scenes with enhanced contextual awareness.

1.3 PhD Thesis Statement/Hypothesis:

"By utilizing natural language processing (NLP) and transformer based language models on on-scene conversational audio data from the care providers and textual corpus from Emergency Medical Services (EMS); and by effectively combining attention-based deep learning techniques and egocentric NLP captioning on smartwatch-based sensor and image data, respectively, it is possible to build intelligent, interactive components of a cognitive assistant for emergency care providers, and thereby moving the state-of-the-art toward more comprehensive and automated EMS training, on-scene and post-scene solutions for the first-responders."

The rest of the thesis is organized as follows. Chapter 2 presents the literature review - related work and state-of-the-art for developing different components and a complete cognitive assistant for EMS. Chapter 3 illustrated by a component called GRACE, for generating summary reports automatically for cognitive assistance in Emergency Response. Chapter 4 describes emsReACT, which is a real-time interactive component of a cognitive assistant for cardiac arrest training in Emergency Medical Services. In chapter 5, we present EMS-BERT - a pre-trained language representation model for the EMS domain. Two of our ongoing research, SenseEMS and EgoCap are discussed in chapter 6, detailing the potential use of smartwatch based sensor data and first-Person image captioning with context fusion for multimodal assistance in Emergency Medical Services.

Chapter 2

Literature Review

In this chapter, we discuss different aspects and features of the state-of-the-art and relevant work for designing and developing cognitive assistants for Emergency Medical Services. We also highlight how our components GRACE (chapter 3), emsReACT (chapter 4), EMS-BERT (chapter 5), SenseEMS and EgoCap (chapter 6) compare against the literature.

2.1 Related Work for GRACE

To the best of our knowledge, GRACE is the first work to address the problem of automatic documentation for the EMS domain. Although, there has been a lot of work on developing smart assistants for emergency response, none of those focus on form-filling. The following subsections review the literature for automated form-filling for emergency situation.

Montanga et al. (Montagna et al. 2019) present TraumaTracker, a trauma tracking system for documentation. They demonstrate that the accuracy of trauma documentation significantly improves after using TraumaTracker, as the system adds data and information that were not recorded in the paper documentation. But this system is deployed only in the trauma domain, GRACE is more generic and can be used for any medical emergency scenario, if the documentation format is similar. Preum et al. in

(Sarah Masud Preum, Shu, Ting, et al. 2018), Shu et al. in (Shu, S. Preum, M Pitchford, et al. 2019) and Lindes et al. in (Lindes, Lonsdale, and Embley 2015) discuss the idea of developing cognitive assistant systems to improve and aid the awareness of first-responders. However, they do not focus on EMS incident report generation or documentation for the patients involved.

Transportation, health and many industry applications have seen different cognitive assistant systems over the years. Authors in (Ha et al. 2014) illustrate a Google glass based assisting system, which is developed to perform context-aware real-time scene interpretation by identifying objects for people suffering from cognitive decline. While the system is useful for this group of people, emergency situations often result in compromising visual capabilities and video signals may not always carry the whole information due to missing angles, and other adverse conditions. Thus, audio data and on-scene conversations are more trustworthy sources for EMS and our module uses them for documentation of patients.

ImageTrend[1] is an increasingly popular tool for documentation, tracking and visualization of EMS information. Another software *Emergency Department Information Exchange (EDIE)*[2] links all hospital emergency departments by facilitating real-time communication and collaboration. However, both ImageTrend and EDIE, require manual input in the initial phase of data collection which is tedious and prone to errors. GRACE does not require any such effort, as summary reports are automatically generated using the audio data from on-scene EMS conversations.

Different tools exist for extracting information from unstructured clinical texts, including, MetaMap (Aronson and Lang 2010), cTAKES (Savova et al. 2010), and

[1] https://www.imagetrend.com/
[2] https://collectivemedical.com/ed-utilization/

CLAMP (Soysal et al. 2017). MetaMap combines natural language processing (NLP) with knowledge-intensive approaches for clinical concept identification and mapping or normalization. The Clinical Text Analysis and Knowledge Extraction System (cTAKES) combines rule-based and machine learning techniques to achieve this. CLAMP is a comprehensive clinical Natural Language Processing (NLP) software that enables recognition and automatic encoding of clinical information in narratives. All three of MetaMap, cTAKES and CLAMP use the Unified Medical Language System (UMLS) to extract medical concepts. One of the main issues with using these tools for EMS documentation or form filling is categorizing the contexts in finer granularity. For example, MetaMap has Concept Unique Identifiers (CUI) and semantic type lists which signify whether a clinical concept is 'Disease' or 'Medication'. But there is no way to differentiate whether the disease or medication is the current condition of the patient or an occurrence from the past. GRACE, on the other hand, uses NLP based heuristics to categorize contexts in finer granularity which is necessary for filling the form. There have been some other works on clinical document summarization and information extraction including (Y. He 2016; Mujjiga et al. 2019). However, they focus only a subset of information relevant for EMS documentation and require significant amount of annotated data, which is not available for the EMS domain.

2.2 Related Work for emsReACT

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 (Koutitas, S. Smith, and Lawrence 2020) 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 (Guo, Fu, et al. 2017) 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 (Guo, Ren, et al. 2016) and (Wu et al. 2014) 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 (Daily, Bishop, Steiner, et al. 2007) suggest that automated, real-time assistants for EMS training will ensure improvement of rescue quality. Authors in (Rahman, S. Preum, et al. 2020; Rahman, Sarah M Preum, et al. 2020) have addressed the challenges discussed in this chapter, 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. The following sections highlight related cognitive assistants from relevant domains, and explain how emsReACT is unique from existing literature.

Lindes et al. in (Lindes, Lonsdale, and Embley 2015) discuss the idea of developing cognitive assistant systems to improve the awareness of EMS providers. Preum et al. (Sarah Masud Preum, Shu, Ting, et al. 2018; S. Preum et al. 2019b) and Shu et al. (Shu, S. Preum, Pitchford, et al. 2019) presented a voice-based cognitive assistant system for suggesting interventions to EMS providers in real-time. Montanga et al. (Montagna et al. 2020) present TraumaTracker, a trauma tracking system for trauma patients. The authors demonstrate that the accuracy of trauma recovery significantly improves after using TraumaTracker, as the system adds data and information that were not recorded in the initial paper documentation. This system mainly emphasizes the documentation aspect. Compared to these systems, emsReACT is more effective in the essence of providing feedback and suggestions in real-time during training. Authors in (Rahman, Sarah M Preum, et al. 2020) and (S. Preum et al. 2019b) provide new methods on top of the state-of-the-art techniques for clinical concept extraction. These methods are modified for real-time use for emsReACT.

Authors in (Sarah Masud Preum, Munir, et al. 2021) provide a detailed survey on types of cognitive assistants in healthcare and other domains. Many cognitive assistants directly interact with a target user in real-time. For instance, authors in (Qian, Deguet, and Kazanzides 2018) present an assistant for robotic surgery which interacts with the human surgeon. However, based on different requirements, often cognitive assistants may interact with multiple users. RoNA (Hu et al. 2011) is a humanoid, mobile robotic nursing assistant for lifting and moving patients and heavy objects inside a hospital to increase patient and nurse safety and operational efficiency. It interacts with an operator through a visual interface where the operator can see and control movements. Similarly, some cognitive assistants interact with the patient and their primary caregiver (Pollack et al. 2002; R. Li, B. Lu, and McDonald-Maier 2015) or professional healthcare provider (Rajanna et al. 2016). emsReACT is unique because it provides feedback through an interface which is accessible to multiple EMS providers in real-time, the feedback mechanism is situation-aware and customized according to different skill levels of the EMS providers.

The modes of interaction for different cognitive assistants are mainly verbal and nonverbal (Sarah Masud Preum, Munir, et al. 2021). A mixed reality (MR) based assistant discussed in (Gamberini et al. 2009) provides training elderly individuals through interactive games in real-time, the authors discuss natural interaction through a tabletop MR platform. Tabletop interfaces mainly use touch-screens and multi-touch technologies, they do not require using a mouse or a touch-pad. A lot of the assistants often perform multi-modal interaction by combining both verbal and nonverbal interactions (Rincon et al. 2019; DeVault et al. 2014). Authors in (Koutitas, K. S. Smith, et al. 2019) proposes augmented reality and virtual reality based technologies for EMS training. However, we consider using the audio data as the safest mode. Visibility issues such as smoke, loss of power may interrupt augmented reality and virtual reality based assistants during training and real EMS scene. Only few of these interaction mechanisms address the challenge of varying degree of users' expertise level. Also, some of the above mentioned operations are not as safety-critical and time-sensitive as EMS training. emsReACT, on the other hand, provides customized interactions in a time-sensitive manner to EMS providers during EMS training. The mode of feedback is textual as audio feedback may interrupt the ongoing process for the EMS providers. However, audio feedback feature is available if the providers choose to use it.

Existing context-aware systems usually have some underlying representation of contexts, these systems learn the context in different ways. The set of contexts for a cognitive assistant can be categorized into temporal, spatial, the user or personal, and situational contexts (Kolenik and Gams 2021). For a real-time cognitive assistant, the user context includes a user's physiological, psychological, behavioral, and medical context. emsReACT provides real-time feedback and they are customized to one or more of the following EMS provider and/or patient contexts.

- Psychological contexts which refer to a human's emotion, mood, personality, level of positivity, and other psychological factors. For example, in EMS the first-responders are prohibited to provide CPR to patients beyond 5 minutes as the provider's psychological factors may degrade afterwards. emsReACT incorporates this constraint in it's design.

- Behavioral contexts which encompass an EMS provider's behavior, action, predefined priority or preferences, level of skills, and professional training and certification. emsReACT uses the level of certification and skills of the EMS providers to understand such context in real-time during EMS training.

- Medical contexts which refer to a patient's current and past medical history, present medical condition, symptoms, diagnosis, medications, genetic profile, family history, and similar medical factors. emsReACT uses dynamic patient condition to understand the medical context during EMS training in real-time.

- The adaptive context or situational context includes environmental context, process context, and operational context. emsReACT is aware of situational context in terms of ongoing process, operations, and predefined protocols. From training scene audio data, emsReACT provides adaptive feedback by assessing the dynamic context in real-time during EMS training for better learning experience for the EMS providers.

2.3 Related Work for EMS-BERT

Different tools exist for extracting information from unstructured clinical texts, including, MetaMap (Aronson and Lang 2010), cTAKES (Savova et al. 2010), CLAMP (Soysal et al. 2017) and EMSContExt (Sarah Masud Preum, Shu, Alemzadeh, et al. 2020). MetaMap combines natural language processing (NLP) with knowledge-intensive approaches for clinical concept recognition and mapping for normalization. The Clinical Text Analysis and Knowledge Extraction System (cTAKES) combines rule-based and machine learning techniques to achieve this. CLAMP is a comprehensive clinical Natural Language Processing (NLP) software that enables recognition and automatic encoding of clinical information in narratives. EMSContExt uses a weakly supervised approach for recognition of EMS concepts from textual corpus leveraging lexical, medical and EMS domain knowledge integration. All of these tools and methods use either the Unified Medical Language System (UMLS), or lexicon expansion approach to extract medical concepts. Two of the main drawbacks of using these tools for EMS entity recognition is their inability for categorizing the contexts in finer granularity, and lack of correlation understanding. For example, MetaMap has Concept Unique Identifiers (CUI) and semantic type lists which signify whether a clinical concept is 'Disease' or 'Medication'. But there is no way to differentiate whether the disease or medication is the current condition of the patient or an occurrence from the past, i.e., recognition of context or relations between entities. Our proposed system EMS-BERT, on the other hand, uses a domain specific bidirectional transformer (BERT) based language model and a simultaneous pre-training technique to recognize entity, their relation and inferring missing information from an unstructured, relatively small-sized EMS corpora. There have been some other works on clinical document summarization and information extraction including, (Devarakonda and Tsou 2015; Y. He 2016; Mujjiga et al. 2019). However, these works focus only on a subset of information and are not specialized for the EMS domain.

The introduction of transformer based language models such as the Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2018) has significantly increased the performance of information extraction from free text. Previously, authors in (Pennington, Socher, and C. D. Manning 2014) proposed a vector representation for words called GloVe embeddings. GloVe does not explore the context while creating the word embeddings which means that the meaning of any specific word in different contexts will render the same embeddings. To address this limitation, authors in (Peters, Neumann, Iyyer, et al. 2018) came up with the idea of contextualized word-embeddings called ELMo, which created word embeddings using a bidirectional LSTM. ELMo is trained with a language modeling objective. ULMFiT (Howard and Ruder 2018) is another successful model for training a neural network with language modeling objective and fine-tuning for a specific task. However, all these models take into account the next occurring words and disregard the context from the previous words. BERT on the other hand, addresses the limitations in these prior works by taking the contexts of both the previous and next words into account instead of just looking at the next set of words for context. BERT (Devlin et al. 2018) is a contextualized word-representation model which is based on masked language modeling (MLM). BERT model is pre-trained using bidirectional transformers (Vaswani et al. 2017). There are two steps in the BERT framework: pre-training and fine-tuning. During pre-training, the model is trained on unlabeled, large-sized corpora. For fine-tuning, the BERT model is first initialized with pre-trained weights, and all the weights are fine-tuned using labeled data from the downstream tasks. BERT pre-training is optimized for two unsupervised classification tasks - masked language modeling (MLM) and next sentence prediction (NSP). The training instance of MLM is a single modified sentence. Each token in the sentence has a 15% chance of being replaced by a special token [MASK]. The chosen token is replaced with 80% of the time, 10% with another random token, and the remaining 10% with the same token. The MLM objective is to find a cross-entropy loss on predicting the masked tokens. Next sentence prediction (NSP) is a binary classification task for predicting whether two segments follow each other in the original text. Positive instances are created by taking consecutive sentences from the text corpus. Negative instances are created by pairing segments from different documents. Positive and negative instances are sampled with equal probability. The NSP objective is designed to improve the performance of downstream tasks, such as natural language inference, which require reasoning regarding the relationships between pairs of sentences. Figure 2.1 (Wada et al. 2020) shows the basic architecture of a BERT model.

Recent BERT models such as RoBERT and ToBERT (Pappagari et al. 2019) provide solutions for classification on long text, KnowBERT (Peters, Neumann, Logan IV, et al. 2019) incorporates different knowledge bases into BERT. In the masked language modeling approach of BERT, words in a sentence are randomly erased and replaced with a special token. A transformer is used to generate a prediction for the masked word based on the unmasked words surrounding it. With the masked

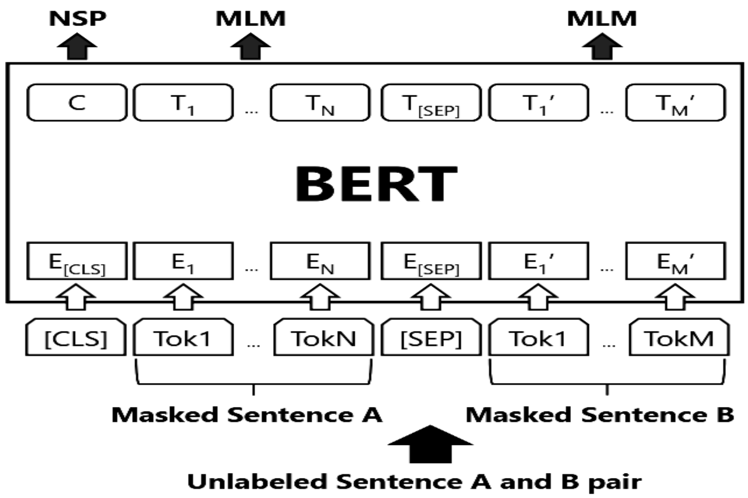


Figure 2.1: Basic architecture of BERT

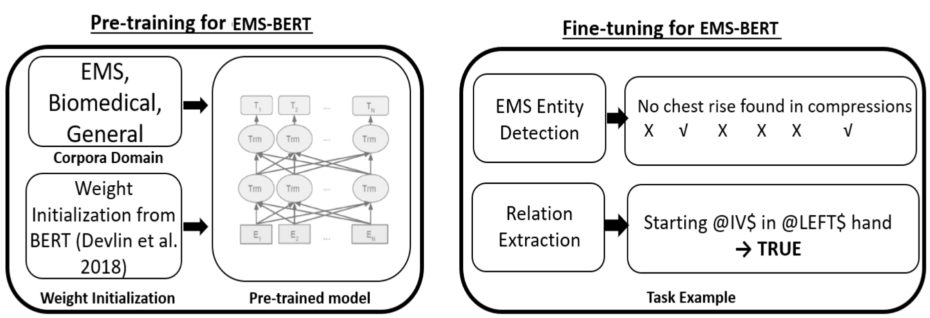


Figure 2.2: Overview of EMS-BERT model

language modeling objective, BERT has achieved improved results for many NLP tasks. Different research with BERT such as BioBERT (Lee et al. 2020), Clinical-BERT (Alsentzer et al. 2019), and SciBERT (Y. Peng, Yan, and Z. Lu 2019) showed that additional pre-training of BERT models on a large domain-specific text corpus results in satisfactory performance in their specific text-mining tasks. For the clinical domain, models such as BioBERT and ClinicalBERT perform text-mining on structured clinical corpora. However, EMS or Emergency Medical Service domain is very different from traditional clinical corpora. The EMS dataset is mostly unstructured, the providers use different sets of semantics and lexicons during their communication and post-incident summary reports. For data mining on our EMS corpora which consists of live-transcriptions and post-scene narratives, existing medical and clinical BERT models do not perform as well as they perform on their respective domain. Hence we develop the EMS-BERT language model for entity recognition, relation extraction and inferring missing information from an EMS corpus.

2.4 Related Work for SenseEMS and EgoCap

Hand gesture detection using smartwatch-based sensor data has gained significant attention in recent years. Numerous studies have explored different approaches and algorithms to accurately detect and recognize hand gestures using data from various sensors embedded in smartwatches. Different studies such as (Kunwar et al. 2022; Wen, Ramos Rojas, and A. K. Dey 2016; Zhu et al. 2018) proposed machine learning-based approaches that utilized accelerometer and gyroscope data from a smartwatch to detect and classify hand gestures. Another research (Bi et al. 2021) focused on combining data from multiple sensors, including accelerometer, gyroscope, to enhance the accuracy of hand gesture detection and present the design, implementation and evaluation of a smartwatch-based, freehand human–computer interaction system. These studies, along with several others highlight the potential of smartwatch-based hand gesture detection in EMS and provide valuable insights into the design of effective algorithms and systems for real-time hand activity detection and related assistance during EMS training and real-scene applications.

In NLP, it is recognized that coherent texts can be summarized through attention (Lebanoff et al. 2020; Zhao et al. 2019). This lends captioning models the power of comprehending the scene using external sources of information. Hence, we are inspired to fuse the contexts as additional information in egocentric way. Conceptual Caption dataset (Sharma et al. 2018) harvests over 3 million image-text description pairs from the Internet, which manifests semantic diversity whereas blurs boundaries between first-person and other narratives. Although popular third-person captioning datasets, such as COCO (Lin et al. 2014), are valuable sources, they cannot be directly used for ego-captioning. Current egocentric visual captioning datasets are limited in either scale or diversity. Charades-Ego (Sigurdsson et al. 2018) and EPIC-Kitchens (Damen et al. 2018) are labelled in Human Activity Classification (HAC) only, and are constrained in terms of scene diversity. Deepdiary (Fan, Zhang, and Crandall 2018) and EDUB-SegDesc (Bolanos et al. 2017) combined release fewer than 300 ego-image samples in total. Ego4D (al. 2021) is a large-scale egocentric video dataset collected across the globe. Unfortunately, the annotations only provide HAC labels and template-based captions like "*A* interacts with *B*". We contrast *EgoCap* with existing datasets with egocentric approach. To summarize, there is currently a lack of sizeable datasets supporting egocentric captioning studies, and this study aims to develop an ego-caption based EMS dataset in future.

Chapter 3

GRACE: Generating Summary Reports Automatically for Cognitive Assistance in Emergency Response

EMS (emergency medical service) plays an important role in saving lives in emergency and accident situations. When first responders, including EMS providers and firefighters, arrive at an incident, they communicate with the patients (if conscious), family members and other witnesses, other first responders, and the command center. The first responders utilize a microphone and headset to support these communications. After the incident, the first responders are required to document the incident by filling out a form. Today, this is performed manually. Manual documentation of patient summary report is time-consuming, tedious, and error-prone. We have addressed these form filling problems by transcribing the audio from the scene, identifying the relevant information from all the conversations, and automatically filling out the form. Informal survey of first responders indicate that this application would be exceedingly helpful to them. Results show that we can fill out a model summary report form with an F1 score as high as 94%, 78%, 96%, and 83% when the data is noise-free audio, noisy audio, noise-free textual narratives, and noisy textual narratives, respectively.

3.1 Problem, Challenges and Overview

Emergency Medical Service (EMS) responders communicate extensively with many different stakeholders in emergency scenarios to ensure that the correct measures are taken and adverse outcomes are minimized. While communicating, the severity of the scene as well as the condition of the injured patients are often mentioned. State-of-the-art technologies such as omni-directional microphones, noise-canceling microphones, headphones, the global positioning system (GPS) and other devices aid the communication and recovery procedure. Currently, a textual narrative of the scene as well as a summary report for the patients are created afterward. These reports often lack critical details that are collected from the scene in real-time, but forgotten. Research shows that in the USA, 13.6% of the time mistakes are made while inputting information into the summary forms. Mistakes happen in the form of inputting wrong information, forgetting to include a correct piece of information and misplacing data in the wrong field of a form (Burnett et al. 2011). Such manual errors can be attributed to the following factors. First, unfavorable circumstances such as getting a call at 2 AM as well as multitasking activities at the scene create adversarial conditions for the first-responders. Second, as responders try to remember the events from the scene, their recall of the events is often not 100% accurate. Finally, most emergency scenes demand dynamic information flow, such as changing vitals, changing medication dosage, etc. which makes the task of post-incident form filling accurately even more difficult. Discussions with first responders indicate that

automatic form filling followed by only needing to check the forms would be a tremendous aid in their jobs.

At first, with the availability of accurate transcription tools and the current state of NLP research, this may seem like a simple task. However, this is not true, as many challenges must be overcome. These challenges include:

1. domain-specific concept extraction that is unique for emergency response when compared to current medical and clinical oriented ontologies, as the specialized vocabulary used by first responders limits the applicability of current solutions;

2. semantic inference from EMS data, e.g., negation detection, temporal expression detection, and value association for accurate information extraction;

3. minimizing the effects of noisy environments and noisy data, missing data, homophones, and other realistic speech issues on information extraction;

4. deep inference of EMS text, including, (a) distinguishing patient-related information from scene and unrelated information in the conversations and (b) chronological ordering of information since the scene is not always narrated linearly.

We developed GRACE (Generating Summary Reports Automatically for Cognitive Assistance in Emergency Response) to solve the above-mentioned challenges. We have collaborated with a regional ambulance agency to get access to 8,000 textual narratives of real EMS scenarios. We also developed 119 simulated audio versions of a subset of the narratives with and without noise to evaluate the variation of the performance of GRACE in presence of noise in speech data, as most emergency scenes are noisy. Further noise insertion in textual corpus is investigated for the validation of GRACE. The main contributions in our chapter are:

- Developed the first NLP based system to address formal documentation or reporting of critical information for emergency response. Our thorough evaluation uses real EMS dataset that includes both textual and speech EMS data. We have explored the applicability of three benchmark NLP clinical information extraction tools for EMS domain, namely, MetaMap (Aronson and Lang 2010), cTAKES (Savova et al. 2010), and CLAMP (Soysal et al. 2017). GRACE outperforms these benchmark tools for information extraction for documentation of emergency response events.

- Demonstrated the impact of noise on audio and textual narratives of emergency incidents and developed a resilient form-filling module that performs acceptably under adverse and noisy conditions. Since emergency response is a low-resource domain in terms of availability of realistic information-rich data, we have generated synthetic noisy conversational data with varying degree and types of noise based on real EMS data for evaluating GRACE.

- Resolved some semantic challenges of domain-specific information extraction for EMS documentation, including, negation detection in EMS text and information validation (e.g., vitals) for EMS data under both noise-free and noisy conditions.

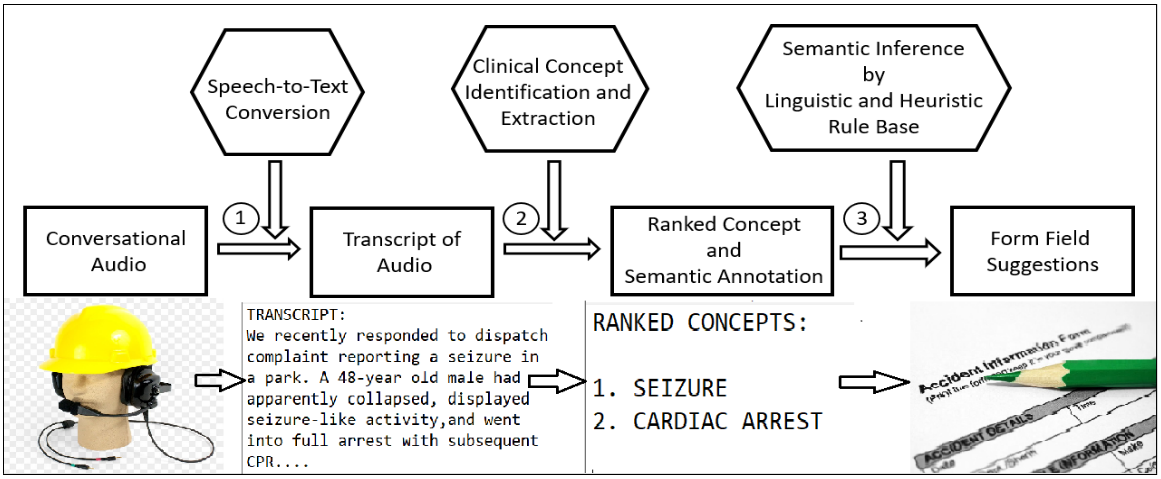


Figure 3.1: Solution steps for GRACE

3.2 Approach and Solution

Figure 4.1 shows the overview of our solution. Although there are many tools to extract medical information, categorizing them into specific fields of the EMS form requires further text processing. This requires additional logic and heuristics compared to the state-of-the-art tools. In the following subsections, we describe our solution.

3.2.1 Speech-to-text conversion

The first step of our solution is speech-to-text conversion, marked by ① in Figure 4.1. There is a lot of noise in EMS scenes, and the accuracy of transcriptions are significantly affected under such adverse conditions (Sarah Masud Preum, Shu, Ting, et al. 2018). Solving this problem is important, but not the subject of this chapter. In the experiments in this chapter, we consider both accurate transcription and noisy transcription to reflect the potential variations in the performance of off-the-shelf speech recognition tools.

3.2.2 EMS concept extraction

After speech-to-text conversion, EMS concepts are extracted from the converted text in step ② (Figure 4.1). Ranking of concepts is done using state-of-the-art medical NLP tools, i.e. MetaMap, CLAMP, and cTAKES. In this chapter, we used Concept Unique Identifiers (CUIs) to filter concepts, and the ranking of concepts is done by using the confidence scores provided by MetaMap. A threshold for confidence score for each type of concept was defined by training our module with training data. Unless a concept ranked above the threshold, it was discarded.

Since MetaMap supports ranking of concepts and unique identifiers according to the confidence scores, it is used in GRACE for clinical context detection and concept extraction. On top of MetaMap, we use different heuristics and linguistic rules to extract necessary information for fields of the form. cTAKES and CLAMP are used for validating the output of MetaMap. First, all the clinical contexts are filtered through MetaMap to discard scene and non-patient related information. We have derived a minimum threshold of confidence score of 5.00 for each of the concepts to be considered. For some of the concepts though, the threshold score is higher. For example, to detect medication and intervention related information, we keep the threshold to 5.00 to ensure all possible concepts are extracted. But for chief complaint or illness history of the patient, our tests with training data illustrate that a threshold of 10.00 works best by omitting false positives. The clinical concepts above certain confidence score are further checked with cTAKES and CLAMP, to ensure that all the state-of-the-art tools identify those as clinical concepts. Unless two of the tools signify a concept as clinical, we discard them. After filtering out non-clinical concepts, we try to understand the semantic meaning (next step below) of each concept and align them with the fields of the form.

3.2.3 Semantic inference

Understanding semantics in textual corpus is a challenging problem and different techniques for identifying semantics exist in the literature (Mujjiga et al. 2019). For semantic inference (step ③ in figure 4.1) such as **negation detection** and **value association** for vitals (i.e. blood glucose levels, Glasgow coma score, respiratory, blood pressure, pulse, peripheral capillary oxygen saturation or SPO2, etc.), we use modifier selection tools, dependency parsers, and entity recognizers. Specifically, NegEx (Chapman et al. 2001) and Stanford dependency parser (Cer et al. 2010) are used for negation detection and StanfordNER (Finkel, Grenager, and C. Manning 2005) is used for associating vitals to their values. However, without punctuation it is quite difficult to understand the context of the narrative. Researchers have identified various methods for adding punctuation in a text corpus (Say and Akman 1996), and recent developments have seen neural network based approaches. Authors in (Tilk and Alumäe 2016) discussed a recurrent bidirectional neural network for missing punctuation. Although this accuracy is not sufficient, we used their online tool to add punctuation into our transcripts as overall performance of GRACE improves afterwards.

3.3 Experimental Setting

Table 4.2 summarizes our datasets. We have generated synthetic data by adding relevant noise profiles to original noise-free audios, however some of our audio data originally had background noise. We have also used textual data from our collaborator, a regional ambulance agency. To train our module, we have randomly selected half of each type of data shown in Table 4.2, while the other half is used for test purposes. The lengths of the audio files vary from one-minute to four-minutes. The artificial noise was added in continuous and discreet mode, and randomly. The amplitude of noise profiles were as high as the amplitude of the original audio, while the minimum amplitude of noise is half of the main audio. For textual narratives, all 32 annotated versions were randomly chosen and consists of minimum 1,000 words and maximum of 5,000 words. Due to limited and constrained resources, and restrictions in collecting live data in real-world EMS scenarios, we consider our dataset to be sufficient for this research. Also, annotating the dataset by professional EMS personnel is a time-consuming and difficult process. However, we are planning to collect more data from real world EMS training scenes and extend our collaboration with various Advanced Life-Support (ALS) EMS providers to enrich our experiments on this research.

3.3.1 Generating synthetic data

We have used the following five types of data for evaluating GRACE:

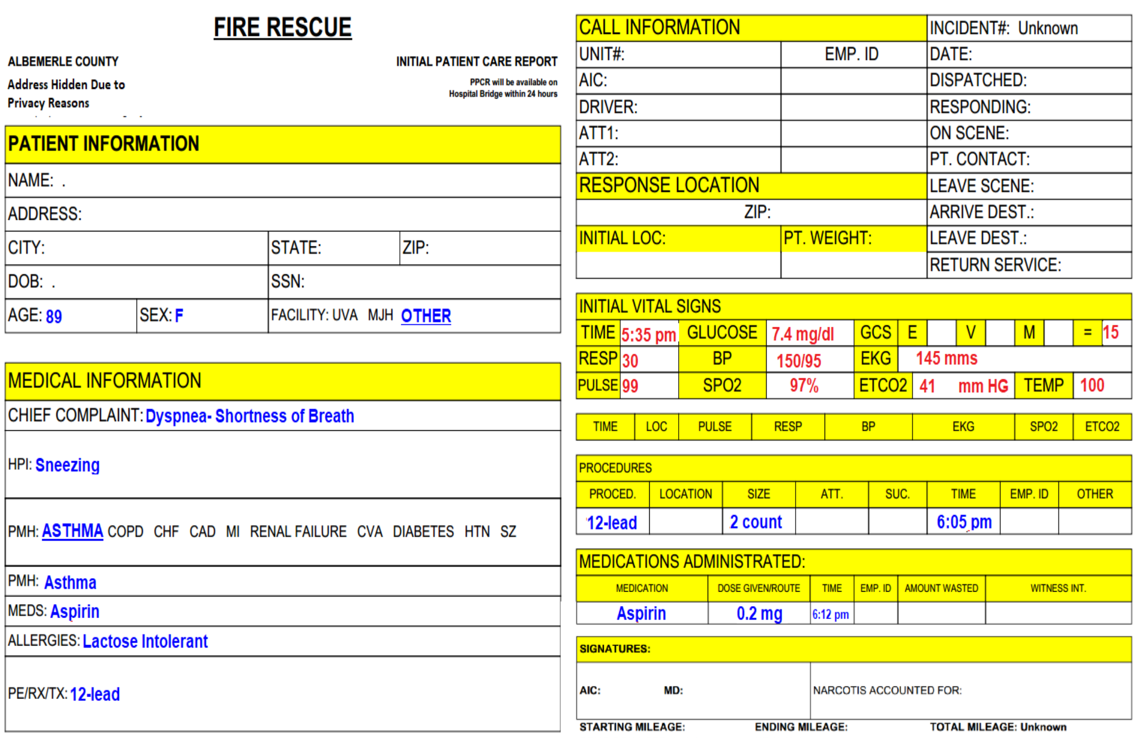
(i) EMS narratives: We have 8,000 post-incident narratives of different EMS scenarios from our regional collaborators. These textual corpora were used to determine the accuracy of our system. Since these data is not annotated and the annotation process is expensive in terms of both time and intellectual effort, this task can not be crowdsourced for reliable and correct annotation. Instead, a small subset of 32 narratives was randomly selected from this data for training and testing purposes.

(ii) Noisy EMS narratives: We utilized different noise-insertion methods in existing research (e.g., (Agarwal et al. 2007)) to insert noise in the textual data mentioned above to validate the robustness of GRACE in presence of textual noise.

(iii) Noise-free audios from the EMS narratives: we have selected 12 test case scenarios from the data we obtained from our regional collaborator (different subset of data when compared to the subset mentioned above) and asked certified EMS responders to simulate a real scene for each of those. There was minimal ambient noise.

(iv) Noisy audio: The same procedure as above was followed, however, there was substantial noise around the scene. The noise was typically people talking in the background, screaming and ambulance siren.

(v) Simulated noisy audio: For the noise-free audio mentioned in the third point above, 8 different types of artificial background noise were inserted with varying degree of intensity. Thus 96 additional synthetic noisy audio data were generated from 12 noise-free audios and 8 noise profiles.



FIRE RESCUE
ALBEMERLE COUNTY
Address Hidden Due to Privacy Reasons
INITIAL PATIENT CARE REPORT
PPCR will be available on Hospital Bridge within 24 hours

PATIENT INFORMATION
NAME: .
ADDRESS:
CITY: | STATE: | ZIP:
DOB: . | SSN:
AGE: 89 | SEX: F | FACILITY: UVA MJH OTHER

MEDICAL INFORMATION
CHIEF COMPLAINT: Dyspnea- Shortness of Breath
HPI: Sneezing
PMH: ASTHMA COPD CHF CAD MI RENAL FAILURE CVA DIABETES HTN SZ
PMH: Asthma
MEDS: Aspirin
ALLERGIES: Lactose Intolerant
PE/RX/TX: 12-lead

CALL INFORMATION | INCIDENT#: Unknown
UNIT#: | EMP. ID | DATE:
AIC: | DISPATCHED:
DRIVER: | RESPONDING:
ATT1: | ON SCENE:
ATT2: | PT. CONTACT:
RESPONSE LOCATION | LEAVE SCENE:
ZIP: | ARRIVE DEST.:
INITIAL LOC: | PT. WEIGHT: | LEAVE DEST.:
RETURN SERVICE:

INITIAL VITAL SIGNS
TIME 5:35 pm | GLUCOSE 7.4 mg/dl | GCS E V M = 15
RESP 30 | BP 150/95 | EKG 145 mms
PULSE 99 | SPO2 97% | ETCO2 41 mm HG | TEMP 100

TIME | LOC | PULSE | RESP | BP | EKG | SPO2 | ETCO2

PROCEDURES
PROCED. | LOCATION | SIZE | ATT. | SUC. | TIME | EMP. ID | OTHER
12-lead | | 2 count | | | 6:05 pm | |

MEDICATIONS ADMINISTRATED:
MEDICATION | DOSE GIVEN/ROUTE | TIME | EMP. ID | AMOUNT WASTED | WITNESS INT.
Aspirin | 0.2 mg | 6:12 pm | | |

SIGNATURES:
AIC: MD: | NARCOTIS ACCOUNTED FOR:
STARTING MILEAGE: ENDING MILEAGE: TOTAL MILEAGE: Unknown
DRUG BOX USED-#. Unknown NEW: Unknown

Figure 3.2: Sample fields in patient summary report form, filled fields colored in blue/red demonstrate the output of GRACE

Table 3.1: Description of synthesized datasets

Type	**Description**	**Size**
Text	EMS narratives	32
	Noise-inserted EMS narratives	32
Audio	Noisy audio (with ambient noise)	11
	Noise-free audio	12
	Audio with artificially injected noise (using 8 noise profiles)	(12*8) = 96

3.3.2 Accuracy metrics

We conducted our experiments according to the form layout from one of our regional collaborators- a local fire response agency. Figure 3.2 shows the fields in the form. The minimum fields required in a post EMS documentation are locally standardized by *ImageTrend* (mentioned in section 2.1), and we included all the required fields in our model report. All the textual and audio data mentioned above was manually annotated according to this form layout by two graduate students working on this project, both of them are certified Emergency Medical Technicians (EMT). The annotations were further reviewed by certified EMS personnel to ensure correctness.Since our target is to measure how accurately GRACE can create summary forms, we have selected *Precision, Recall, and F1 Score* as our accuracy metrics.

3.4 Evaluation

GRACE outputs acceptable accuracy numbers for all the fields in the form shown in Figure 3.2. Typical fields such as Medication Administration, Vital Signs, and

Procedures yield an average F1 score of 0.79, 0.86 and 0.71 for test data that includes noise-free audio, noisy audio, noise-free narratives and noisy narratives. The performances of the medical concept extraction tools (e.g. MetaMap, cTAKES, and CLAMP) are also comparable for these information fields. Due to limitation of space, we omit further details for these fields. Also, because of the chapter space limits and because of their difficulties and importance we choose to show results of our negation detection step and the filling in the most important fields, including Chief Complaint, HPI (History of Present Illness), and PMH (Past Medical History). Appendix A holds detailed results for all the fields.

3.4.1 Performance of negation detection

Although state-of-the-art tools use an enriched set of rule bases for detecting negation in clinical texts and electronic health records, it is difficult to identify if sentences have multiple negations or multiple contexts. For example, transcriptions such as *patient denied having shortness of breath* or *patient denied having lack of chest pain* contains double negations. Multiple negated contexts are also difficult to determine, e.g., "*patient denied having headache, shortness of breath and chest pain*". Another issue is that ill-punctuated transcriptions create lots of false positives in our train and test data while detecting negation. We have experimented with off-the-shelf and state-of-the-art tools such as DEEPEN (Mehrabi et al. 2015), NegEx (Chapman et al. 2001), MetaMap, and cTAKES. The accuracy of each tool is shown in Figure 3.3, the experiment was done with all of the test data in Table 4.2. NegEx outperforms all the other tools; the F1 score of NegEx is 0.81 with the highest precision compared to other tools. Although recall of DEEPEN is higher than NegEx (0.77 compared to 0.76 of NegEx), but the precision and F1 score is lower for our data. MetaMap and cTAKES have built-in negation detectors which can be used solely for detecting negative phrases, but they perform poorly; their F1 score is 0.44 and 0.49, respectively. Since NegEx performs the best, we adapt NegEx for GRACE. Additional customization is done on top of NegEx by adding to the existing rule base and incorporating heuristics for detecting multiple negated contexts in a sentence.

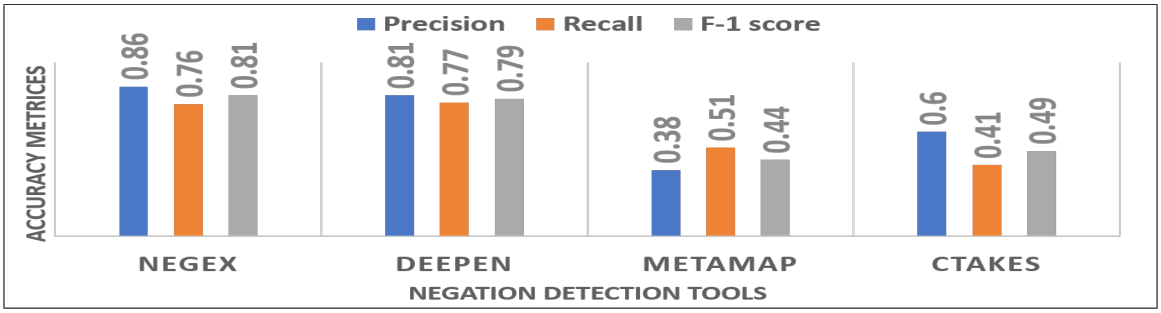


Figure 3.3: Baseline comparison for negation detection

3.4.2 Accuracy of critical medical information

Chief Complaint

The chief complaint (CC) of the patient is challenging to define as there are multiple clinical contexts in the narrative. Medical information extraction tools provide different tags for chief complaint, e.g. "sign or symptom" by cTAKES; "findings", "sign or symptom" and "injury or poisoning" by MetaMap; or"problem" class in CLAMP. But these tags could relate to any of the contexts of other fields in the form also, such as past medical history, history of present illness, allergies and so on. On top of the tools used, hypothesis developed in GRACE detects the most likely candidate for chief complaint from the contexts in the transcription. Figure 3.4 summarizes the accuracy of our findings for chief complaint, and also demonstrates the comparison with the state-of-the-art tools. For the clarity of the figure and due to space limitations, we show the F-1 scores only. The tags mentioned above were used for each of the tools to extract the chief complaint candidates. These results are compared with the ground-truth data annotated by a real EMS responder.

We apply different heuristics and keyword identification for determining the chief complaint of the patient. Our investigation with EMS transcripts reveal that the chief complaint of the patient is generally mentioned at the beginning of scene description. Lack of correct punctuation causes difficulty in understanding the semantic meaning, thus we apply the punctuation insertion mechanism discussed in (Tilk and Alumäe 2016), after the speech-to-text conversion step. The resultant narratives are filtered for clinical concepts by MetaMap, cTAKES and CLAMP; we only select the clinical concepts that are found up to first three sentences. These clinical concepts have higher probability of holding the information of chief complaint of the patient. GRACE seeks for any mention of phrases like *"The patient is complaining of"* or *"Chief complaint is"* or *"The patient is suffering from"*, and if found then finds which clinical concept(s) are related to that phrase using dependency parsers. The first two clinical concepts with highest confidence scores (determined by MetaMap, cTAKES and CLAMP) are selected as chief complaints unless such phrases are mentioned explicitly. If at least one common concept does not exist in the output of all three tools, we leave the field empty for post-scene manual input by first-responders with a highlighted remark to draw their attention.

The implication of the result in Figure 3.4 is two-fold. First, state-of-the-art tools are far off from defining clinical information in finer granularity. Although the concepts in concern are detected fairly accurately (acceptable precision), but the false positives and false negatives are too high (poor recall). MetaMap, CLAMP and cTAKES has an average F1 score of 0.65, 0.64 and 0.62 respectively for noise-free audio, noisy

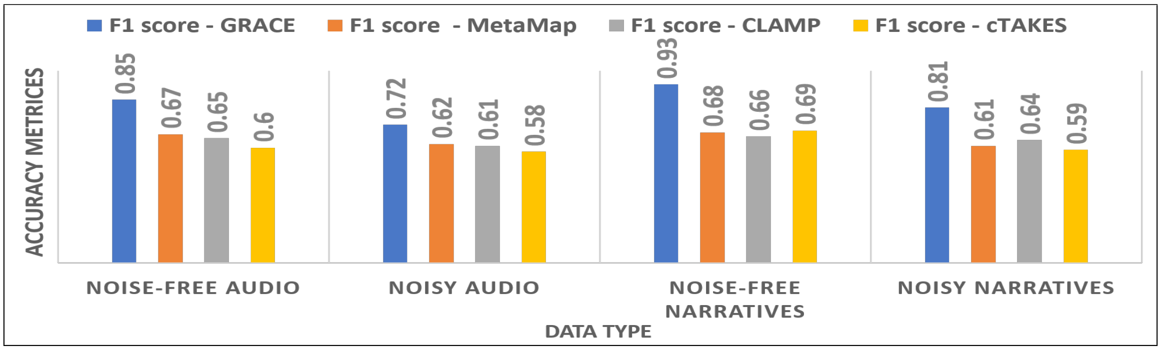


Figure 3.4: GRACE outperforms the state-of-the-art clinical information extraction tools for detecting Chief Complaint from each EMS dataset used in our evaluation.

audio, noise-free narratives and noisy narratives while determining chief complaints. GRACE uses additional logic and filter to narrow down the possible results and achieves an average F1 score of 0.83. Second, many of the concepts are previous symptoms or past history, but they were detected as probable chief complaint by the tools. When using only the tags mentioned above, the tools return a lot of clinical concepts, most of which are effects of the chief complaint or related to the development of current conditions of the patient. GRACE, on the other hand, uses additional heuristics, ranking, and semantic inference to distinguish the clinical concepts, and selects chief complaint with better accuracy. Our understanding is that assuming the chronological development of patient's clinical condition in the transcription plays an important role in increased F1 score of GRACE, 0.85, 0.72, 0.93 and 0.81 for noise-free audio, noisy audio, noise-free narratives and noisy narratives, respectively. Most of the information transcribed in the middle or later sections of the audio data do not contribute to the chief complaint; information in the beginning holds all the true positives.

History of Present Illness (HPI) and Past Medical History (PMH)

History of present illness (HPI) and past medical history (PMH) are very important information to understand patients' condition and the development of the symptom. Empirically, there are certain keywords and phrases which first-responders use to signify HPI and PMH, for example *"Patient has been feeling stomach ache for two days"* or *"She took pregnancy-related pills two months ago"*. Our heuristics use state-of-the-art NLP tools to understand the difference, and determine possible candidates for both of these fields in the form. The significance of detecting correct information in these two fields are very important, as the range of candidates span from clinical concepts to daily activities which might be linked with the current condition of the patient. Past information related to allergies are also critical, because many of our false positives are caused due to miss-classification of this information, and interchanged content in these fields. Our heuristics only rely on specific keywords for this part, however we use an entity recognizer and different NLTK classifiers to separate the related information. Figure 3.5 and 3.6 summarize our findings for HPI and PMH respectively, comparison with other tools is irrelevant as there is no specific tag or semantics provided by these tools to identify the two categories. One important thing to mention here is that our module is tested on data which do not have any time-stamps, we assume that chronological ordering of development of patient's symptoms is maintained while transcribing. Average F1 scores for HPI and PMH are 0.71 and 0.70. This is due to the inability of GRACE to understand the context at times due to lack of proper punctuation and noise in transcriptions. Within sentence boundaries, transcribing multiple symptoms which relate to different fields of the form adds to the challenge. One important thing to mention for all fields of the form is that no specific keyword or verbalization was predefined while generating the synthetic data. It is our understanding that explicit mentioning of the context and better noise-canceling techniques can improve the accuracy of these fields.

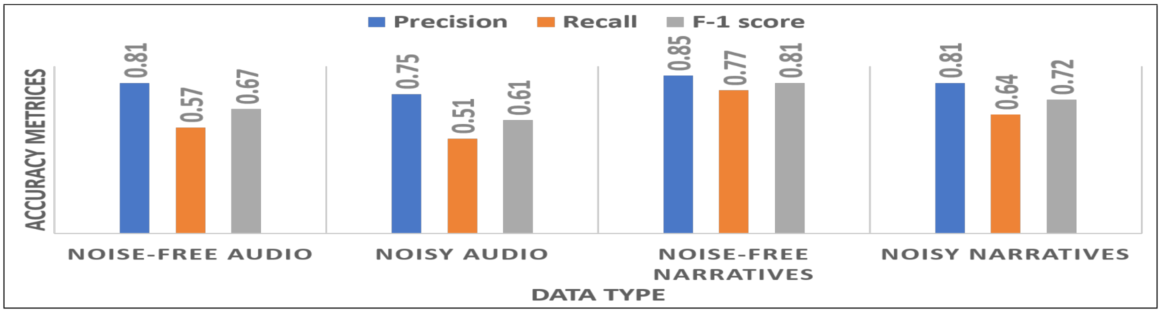


Figure 3.5: Accuracy of GRACE for detecting HPI

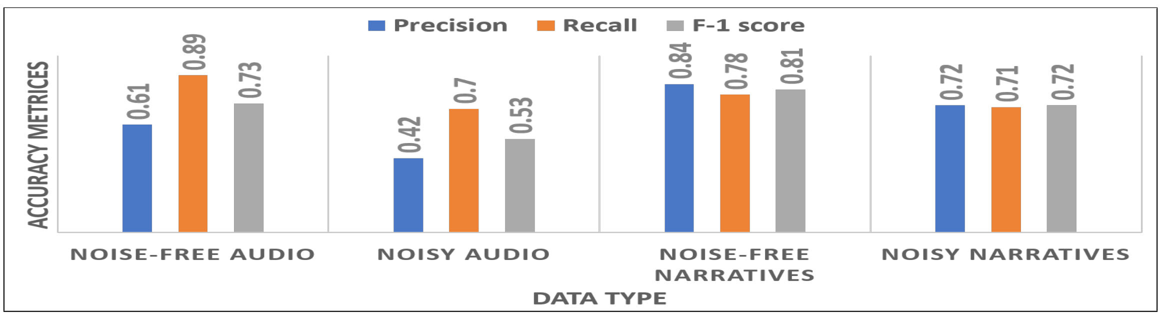


Figure 3.6: Accuracy of GRACE for detecting PMH

3.5 Conclusion

This chapter addresses the problem of automatic summary report generation for patients involved in an EMS scenario. Using simulated audio data from the scene and conversations from first-responders, we show that our solution can generate an initial summary report by filtering and identifying relevant EMS information and context. We are the first to show that such documentation can be done with an F1 score as high as 94%, 78%, 96%, and 83% when the data is noise-free audio, noisy audio, noise-free textual narratives, and noisy textual narratives, respectively. Due to approval issues, we are yet to test our system in real-world EMS scenarios, but we are planning to deploy the system in EMS training soon. Our solution is not robust to all kinds of error and noise at the moment; however we claim that GRACE is very helpful for first-responders as it provides them with an initial draft of the summary of an injured patient, which can further be modified later manually if needed, to finalize post EMS scene documentation. The EMS responders do not have to completely depend on their memory for the task; and even though the accuracy is not perfect, the first-responders will highly benefit by the automate initial draft. In the future we plan to highlight missing interventions and critical inconsistencies detected from the conversation regarding patient's clinical condition. We also aim to develop a more generic and scalable approach by considering multi-patient and multi-responder scenes, and by applying machine learning techniques.

Chapter 4

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

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.

4.1 Problem, Challenges and Overview

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 (Graham, McCoy, Schultz, et al. 2015). 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 (Sarah Masud Preum, Munir, et al. 2021), 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 (Burnett et al. 2011). 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 (Aronson and Lang 2010), cTAKES (Savova et al. 2010), and CLAMP (Soysal et al. 2017) 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.

4.2 Background on Cardiac Arrest

This section presents a brief background on cardiac arrest protocols, necessary interventions and their related EMS training procedures.

4.2.1 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) (Parish, Goyal, and Dane 2018). In this chapter, 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 4.2. 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.

4.2.2 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. EMT-advanced, EMT-enhanced and paramedics certification holders are ALS providers. Table 4.1 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 4.1: 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
assist ventilation (bvm)	EMT-Basic	2	4	1	Check Pt allergies
cardiac monitor	EMT-Basic	1	4	1	Check Pt allergies
chest decompression	Paramedic	1	4	4	Check Pt allergies
CPR	EMT-Basic	4	4	2	Allergies
defibrillation	EMT-Basic	4	4	3	Allergies
intubation	EMT-Advanced	4	4	4	Allergies
oropharyngeal airway insertion	EMT-Basic	1	4	2	Check Pt allergies

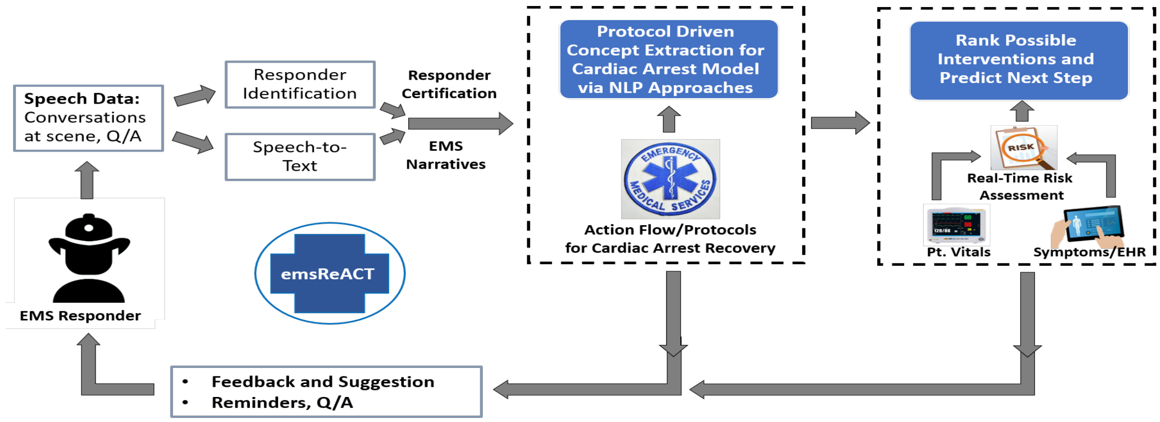


Figure 4.1: emsReACT solution overview

4.3 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 4.1 shows the high-level architecture of the system. The following subsections 4.3.1,4.3.2, and 4.3.5 briefly describe the overall assistant and are included for completeness. The subsections 4.3.3 and 4.3.4 detail the main contributions of real-time dynamic scheduling for this chapter.

4.3.1 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 (Sarah Masud Preum, Shu, Ting, et al. 2018). In the experiments of this chapter, 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 chapter, we do not detail the process here. We use the state-of-the-art Google Speech API for this step.

4.3.2 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 4.1. For extraction of concepts from the text, state-of-the-art clinical NLP tools, i.e. MetaMap (Aronson and Lang 2010), cTAKES (Savova et al. 2010), EMSContExt (Sarah Masud Preum, Shu, Alemzadeh, et al. 2020) and CLAMP (Soysal et al. 2017) 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 (described in Chapter 5) for detecting concepts from the speech using a lexicon expansion approach. We developed a detailed cardiac ontology (**müller2006sudden**; Narayan, P. J. Wang, and Daubert 2019) 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 this contribution.

4.3.3 Task abstraction for scheduling an anytime feedback

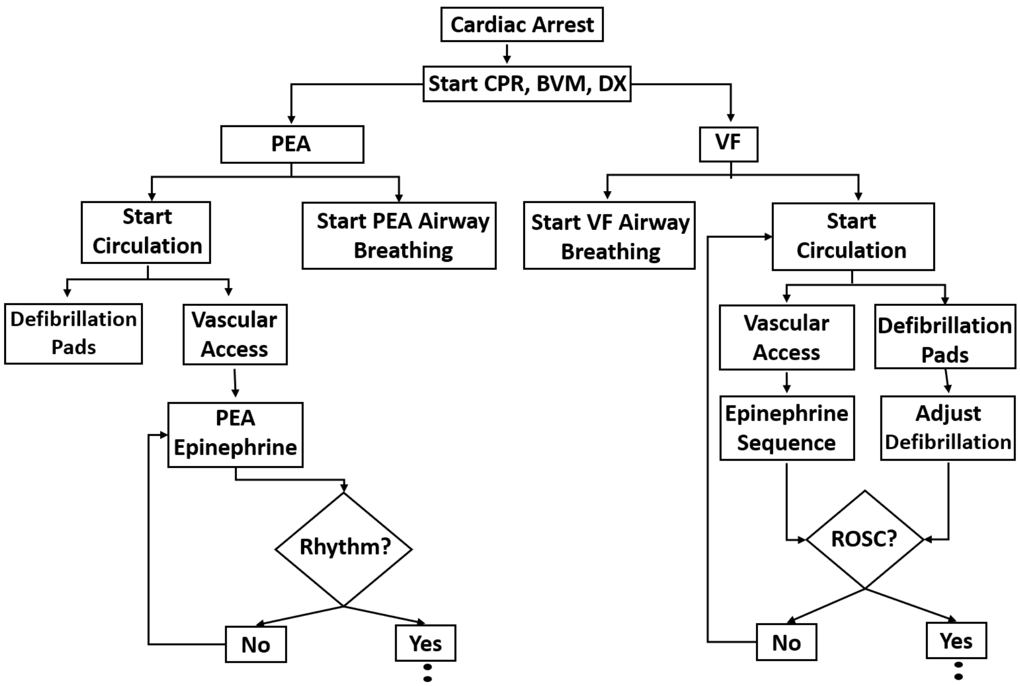


Figure 4.2: Intervention flow (partial) for VF and PEA recovery

Cardiac arrest protocols do not follow any static flow of action, rather the overall procedure consists of many different dynamic actions or tasks (Figure 4.2). 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

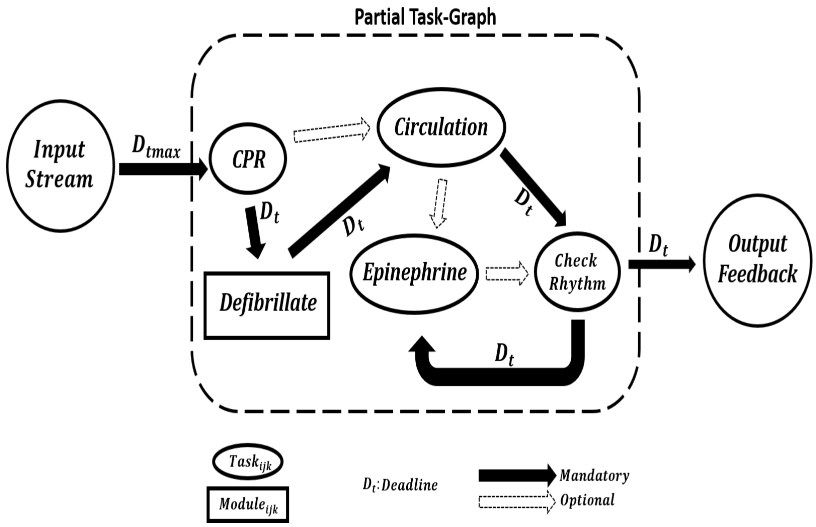


Figure 4.3: Task abstraction concept for emsReACT

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 (Yao et al. 2020). 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.

A key component of our solution is creating the task graph. The entire patient recovery process from Figure 4.2 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 4.3 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 4.3.4). 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.

4.3.4 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 an anytime algorithm (Algorithm 1 in Figure 4.7). 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 4.1 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 4.3.3, 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.

4.3.5 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

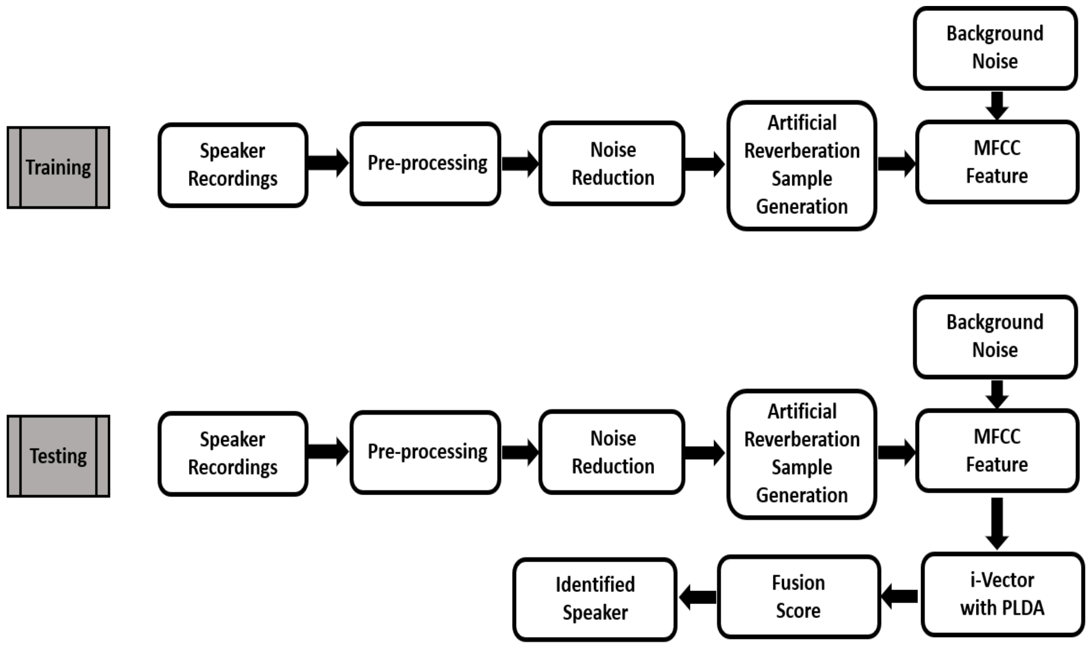


Figure 4.4: Speaker identification technique for emsReACT

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 such as(Seki et al. 2018), however the training phase is costly. Deep neural network based solutions (W. He, Motlicek, and Odobez 2018) are not effective for real-time EMS environment. We apply the basic method proposed in ARASID (Z. Chen et al. 2019), 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 as this is out of scope for the contribution of this chapter.

Figure 4.4 shows the architecture, training and testing details of ARASID for emsReACT. For emsReACT, we generate the MFCC features while training the system with speakers as an artificial reverberation sample generation requires small amount of sample data for training. For testing, we modify the basic ARASID model and generate i-vectors from the features, instead of generating GMM-UBM based speaker identification system. The latter approach requires using different reflection coefficients to model realistic levels of reverb, which is inappropriate for real-time operation. We calculate the fusion score from all the sample and rank the most probable candidate speaker.

4.4 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 (Rahman, Sarah M Preum, et al. 2020) 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 (Sarah Masud Preum, Shu, Alemzadeh, et al. 2020). 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.

4.4.1 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 4.2 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 (Rahman, Sarah M Preum, et al. 2020). However, some of our

Table 4.2: 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

audio data originally had background noise. We have also used textual data from our regional collaborator, a regional ambulance agency (RAA). 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 4.2, whereas the other half is used for test purposes.

4.4.2 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

Table 4.3: 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			

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 4.3 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.

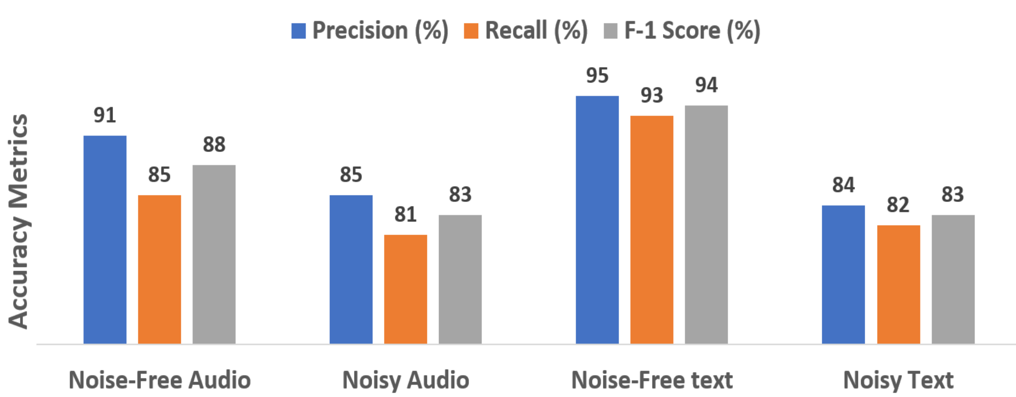


Figure 4.5: Accuracy of emsReACT for different types of data

Performance of Speech-to-Text Conversion under Noise

We evaluate comparative performances of state-of-the-art speech-to-text conversion tools under noise. Three off-the-shelf speech to text APIs - Google speech API, Microsoft speech API, IBM BlueMix API and one offline tool DeepSpeech (Hannun et al. 2014) are compared for accuracy and latency metrics using live speech and audio data in both noisy and noise-free environments, as depicted in Table 4.4. The performance is measured in terms of runtime (seconds) and word error rate (WER) (Sarah Masud Preum, Shu, Ting, et al. 2018). The runtime indicates time needed to transcribe each sentence on average, the word error rate (WER) is indicative of how much noise or distortion exists in the transcription. The Google cloud API outperforms the other tools in terms of WER (at least 16% lower than any of the three tools compared), thus we select this API for emsReACT even though this requires internet connectivity and slightly longer (0.15 seconds) runtime than the offline tool DeepSpeech. Even though the WERs are somewhat high, our solutions are robust to this amount of WER as we use a vocabulary which is comprised of empirical mappings of homophones and distorted versions to original concepts.

Table 4.4: Comparison for training scene transcription tools.

Metric/Tools	Google Speech	Microsoft Speech	IBM BlueMix	DeepSpeech
WER (%)	31	47	49	61
Runtime (s)	0.94	1.08	1.23	0.79

Table 4.5: Comparison for cardiac concept detection and/or generating feedback

Method	Avg. time (s)	Metric	Detecting cardiac concepts	Generalized feedback	Personalized 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

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 (Rahman, S. Preum, et al. 2020) 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 4.5 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 4.5. 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, emsRe-

ACT outperforms IMACS by at least 8% in Precision and by 7% in Recall. Training with a larger dataset increases the accuracy of our solution.

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 4.2. The other half is used for testing. The test dataset shows that for different types of data, average F-1 score is 87% (Figure 4.5) 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.

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 eval-

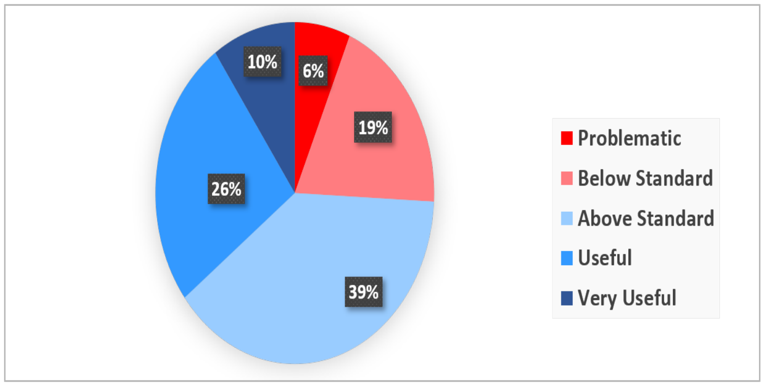


Figure 4.6: Survey from 31 anonymous EMS providers

uation. 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 4.6. 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. 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. Appendix B holds more survey details and an example scene.

4.5 Discussion

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. Overall, the results are promising for training EMS responders for cardiac cases. This is based on our initial survey results that suggest that emsReACT will be influential to affect the training on making decisions in critical situations. In the future we expect that emsReACT can also be used in actual EMS scenes. But further user studies are required to improve the performance of emsReACT where no instructors are present. A dynamic risk-rating based variant of anytime algorithm is used for emsReACT in the EMS domain with acceptable results. We believe that the key solutions developed here, including dynamic understanding of safety critical scenes that have unique vocabularies, dynamically computing safety critical risk indexes, and having such indexes drive real-time anytime algorithms can be used in other applications such as training police and fire, various military training, emergency departments in hospitals, and for in-hospital surgeries. The methods discussed in this research can be extended to address in-home emergency situations using existing systems such as Alexa, Google Home, etc. Our future goal includes using reinforcement learning instead of rule-based solutions for real-time assistance via safety-critical applications.

4.6 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. 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.

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 ← 0
interventionRisk ← 0
ratingRisk ← conceptRisk + interventionRisk
situationVar ← 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 ≥ 7 then
            situationVar ← 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 ← conceptRisk + interventionRisk
    if ratingRisk ≥ 7 then
        situationVar ← Critical
        Output Urgent Feedback (within 3 s)
    else if ratingRisk < 7ANDratingRisk > 0 then
        situationVar ← Regular
        Output Feedback (within 7 s)
```

Figure 4.7: Assessing situational awareness by dynamic risk-rating calculation via a form of an end-to-end Anytime Algorithm

Chapter 5

EMS-BERT: A Pre-Trained Language Representation Model for the Emergency Medical Services (EMS) Domain

Emergency Medical Services (EMS) is an important domain of healthcare. First responders save millions of lives per year. Machine learning and sensing technologies are actively being developed to support first responders in their EMS activities. However, there are significant challenges to overcome in developing these new solutions. One of the main challenges is the limitations of existing methods for EMS text mining, and developing a highly accurate language model for the EMS domain. Several important Bidirectional Encoder Representations from Transformer (BERT) models for medical domains, i.e., BioBERT and ClinicalBERT, have significantly influenced biomedical text mining tasks. But extracting information from the EMS domain is a separate challenge due to the uniqueness of the EMS domain, and the significant scarcity of a high-quality EMS corpus. In this research, we propose EMS-BERT - a BERT model specifically developed for EMS text-mining tasks. For data augmentation on our small, classified EMS corpus which consists of nearly 2.4M words, we use a simultaneous pre-training method for transfer-learning relevant information from medical, bio-medical, and clinical domains; and train a high-performance BERT model. Our thorough evaluation shows at least 2% to as much as 11% improvement of F-1 scores for EMS-BERT on different classification tasks, i.e., entity recognition, relation extraction, and inferring missing information when compared both with existing state-of-the-art clinical entity recognition tools, and with various medical BERT models.

5.1 Problem, Challenges and Overview

Emergency Medical Services (EMS) provide emergency medical care to patients who are involved in an incident that causes serious illness or injury. EMS play an intricate role in healthcare, each component of EMS performs coordinated efforts for providing emergency medical care to the patient(s). An EMS system does not exist in isolation, rather it is integrated with other healthcare related services intended to maintain and enhance a community's health and safety. Emergency services often provide the most timely initial care to begin the recovery process for the patient. In the USA alone, EMS providers save thousands of lives everyday and initiate a primary phase of patient recovery through the healthcare system (Al Amiry and Maguire 2021). To improve healthcare and provide better services to the patients, the EMS domain can not be ignored. Sometimes, the whole recovery phase of the patient is conducted by EMS. EMS providers perform different interventions on the patient, and collect lots of data during an EMS scene for future treatment. The information regarding the recovery process and patient health are documented after each EMS episode. Using this data from the EMS scene, novel methods such as an EMS specific language model can be built to analyze patient information and predict patient outcome. State-of-the-art assistants for medical care heavily rely on correct detection of EMS related medical information. An EMS specific language model can also be utilized in EMS based applications such as (S. Preum et al. 2019a; Sarah Masud Preum, Shu, Ting, et al. 2018; Rahman, Sarah M Preum, et al. 2020) for developing a better, robust and more automated healthcare system.

EMS reports hold significant data related to different EMS protocols, interventions, and clinical conditions of the patient (Rahman, Sarah M Preum, et al. 2020). As EMS scenes occur frequently (especially during the COVID-19 pandemic) (Al Amiry and Maguire 2021), and these reports are always generated afterwards, analysis of such information can also play an important role in optimizing the entire process, i.e., save money, time, and lives by better understanding of the EMS information and their correlation to improve performance in the future (Kim et al. 2021).

Different machine learning techniques exist in the literature to uncover patterns and improve predictions (Yu, Beam, and Kohane 2018; Nguyen-Duc et al. 2021; Xiao, E. Choi, and Sun 2018). However, unstructured, high-dimensional, and sparse information such as EMS reports are difficult to use in traditional machine learning models. In recent years, advances in deep learning and transformers have led to great progress towards generic and personalized predictions in different medical domains. A key contributing factor to this success is the introduction of large multimodal health data such as electronic health records (EHR) (Shickel et al. 2017). Each individual's EHR can link data from many sources, i.e. doctor visits and hospital episodes. This data contains entities/concepts such as diagnoses, interventions, lab tests, clinical narratives, and more. The adoption of EHR systems has greatly impacted the frequency of hospitalization of patients (K. Huang, Altosaar, and Ranganath 2019; Y. Li et al.

2020) and detection of severe illnesses (Poplin et al. 2018; Ardila et al. 2019). On the contrary, the EMS domain has seen almost no advancements in processing the EMS data for understanding the patient condition and personalized treatment generation. Just like the EHR data, an EMS dataset can be utilized to develop a domain specific language model for text-mining purposes in EMS based applications.

For developing a domain-specific language model, pre-training of the language model on large-scale raw textual corpus has already made a tremendous contribution for transfer learning in natural language processing (NLP). Introduction of transformer-based language models, such as Bidirectional Encoder Representations from Transformers (BERT) has significantly improved the performance of information extraction from free text in the general domain (Devlin et al. 2018). For domain-specific purposes, many studies showed that additional pre-training of the BERT model on a domain-specific corpus results in better performance in their specific text-mining tasks. Two of the most contributing factors for developing a domain adapted language model are the size of the training corpus, and the relevance of training dataset. For example, BERT models such as BioBERT and ClinicalBERT localize on biomedical and clinical text, respectively (Lee et al. 2020; Alsentzer et al. 2019). However, these models are developed for the medical domain, and the EMS domain is significantly separate from both. Although the performance of the medical BERT models is good for entity and relation detection tasks, other barriers exist to relate the localization to the EMS domains. For example, information extraction and correlation detection for the EMS domain is unique from the previous two domains when compared to the lexicon of in-hospital medical and clinical corpora (Kim et al. 2021). The EMS domain has its own uniqueness because of the specialized vocabulary which the first responders use. These are the main reasons which limit the applicability of current medical and clinical solutions. Compared to existing clinical and medical dataset, the EMS dataset is often specialized, unstructured and noisy. Our experiments suggest that concept detection as well as semantic inference from EMS data, i.e., negation detection, temporal expression detection, and value association for accurate information extraction requires different approaches compared to the clinical state-of-the-art methods and tools (S. Preum et al. 2019a). However, due to the lack of available datasets in the EMS domain, we devise a solution to utilize the overlapping portion of clinical and medical datasets for augmenting our experimental EMS dataset. Since both domains are based on medical issues, there also exists overlap between EMS and the clinical domain. Some portions of EMS concepts are similar to clinical and medical concepts such as disease names and medication names.

As our EMS dataset is limited, we utilize data augmentation from related clinical and medical BERT models to develop **EMS-BERT: A Pre-Trained Language Representation Model for Emergency Medical Services (EMS) Domain** for text-mining purposes in EMS. For developing EMS-BERT, we implement simultaneous pre-training (Wada et al. 2020) method using two relevant types of corpora and combine them to create a sizable corpus of over 1.5B words. We augment our training corpus with amplified vocabulary from these related domains as well as from the general domain. First, we show the efficiency of our method for downstream text-mining tasks, i.e., entity/concept recognition, relation extraction, and inferring missing information using comparison with predefined EMS protocols on EMS documents. Then, we also demonstrate that when applied on the EMS domain, our approach provides a better pre-trained model that outperforms existing BERT models from the general, medical, clinical, and bio-medical domains, i.e., BERT-Base, BioBERT, and Clinical-BERT, and existing clinical concept recognition methods, i.e., MetaMap, CLAMP, and cTAKES. The main contributions of this research are:

- To the best of our knowledge, EMS-BERT is the first BERT based language model for clinical entity/concept related text-mining purposes on an EMS specific textual corpus. The novelty lies in the new EMS domain where we apply the BERT technique, and in creating the training instances for the model.

- Leveraging the state-of-the-art method of simultaneous pre-training, EMS-BERT is developed using an EMS corpus of real-scene transcripts and post-scene summary reports. This corpus is larger than any other EMS corpus available; and the corpus is augmented with medical, clinical, bio-medical and general corpora with the simultaneous pre-training method. Subsequently, we show that the localization of EMS entities with EMS-BERT is feasible using our method.

- We compared the performance of EMS-BERT with state-of-the-art techniques and without the simultaneous pre-training method. EMS-BERT with simultaneous pre-training outperforms existing BERT based models from relevant domains (i.e. BERT-Base, BioBERT, ClinicalBERT) and existing clinical concept recognition methods (i.e. MetaMap, CLAMP, cTAKES, EMSContExt) in F-1 scores for EMS entity recognition and relation extraction from EMS corpora by 5% to 11%, and by 2% to 6%, respectively. Also, using standard EMS protocols and guidelines, EMS-BERT infers missing EMS information from EMS documents with at least 7% to as much as 14% better accuracy compared to other methods.

5.2 Methodology and Solution

In the following subsections, we describe the underlying methods for EMS-BERT. Figure 4.1 (Wada et al. 2020) shows an overview of the overall approach.

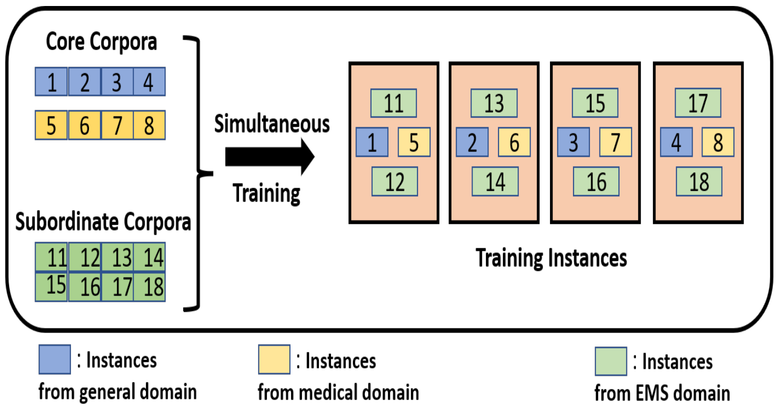


Figure 5.1: Simultaneous pre-training for EMS-BERT

5.2.1 Data augmentation

Through regional collaborators, we have access to 40,000 EMS narratives from real EMS scenes. These EMS narratives constitute over 2.4M words which contain different EMS concepts or entities, i.e., signs and symptoms, interventions, and medication information. Even though we have 40,000 narratives, this is relatively small compared to other dataset sizes for different domain specific BERT models (5.1). Besides the size, most of this narrative corpus are structured as they are created as post-scene summary reports. However, EMS narratives are also created with unstructured, on-scene communication based transcripts. EMS-BERT should be able to extract critical information from on-scene EMS narratives which is created by speech transcriptions collected at emergency scenes. So we need an EMS dataset which contains attributes such as noise, sparsity and of an unstructured nature for training EMS-BERT. For noisy and distorted EMS entities, we created a mapping of distorted entities to original EMS entities for the EMS-BERT model. The tokenization uses this mapping for suggesting the potential correct entity. Besides the noise, to train EMS-BERT with a sizable corpus for including broader medical and clinical entities, we devise a simultaneous pre-training method using a corpus from relevant domains. We have augmented the dataset by the following two methods to include both of these features.

Textual noise insertion.

We utilized different noise insertion methods in textual corpora to emulate EMS narratives created from on-scene transcripts. Since the speech-to-text conversion sometimes yields distortion and inappropriate homophones in the presence of noise, we have used the state-of-the-art noise insertion methods to simulate similar kinds of errors. These noisy textual narratives are used for training EMS-BERT to mimic on-scene EMS transcripts. Authors in (Rahman, Sarah M Preum, et al. 2020) discuss the possible kinds of noise found in textual data. The authors in (Subramaniam et al. 2009) highlight on text produced by processing signals and demonstrate that they are often noisy for automated processing. We have implemented a modified version of SpellMess (Subramaniam et al. 2009) to introduce spelling errors in the EMS corpus. This modified version can change and/or substitute phonetically similar segments in a word, e.g., replacing a word with a homophone. We have created a list of possible homophones found in clinical context from (LaFleur-Brooks and LaFleur 2005). Besides insertion, deletion and substitution of letters, homophone substitution is highly correlated with the kind of impact noise found in EMS transcriptions.

5.2.2 Simultaneous pre-training of EMS-BERT.

The standard BERT model does not perform well in specialized domains (Lee et al. 2020). To overcome this limitation, possible techniques include additional pre-training on domain-specific corpora from an existing pre-trained BERT model, or pre-training from scratch on domain-specific corpora. A main benefit of the former is that the computational cost of pre-training is lower than the latter. The main advantage of the latter is the availability of its custom vocabulary, but the disadvantage is that the pre-trained neural language model may be less adaptable if the number of documents in a specific domain is small. Due to the scarcity of public EMS corpora, both approaches seem infeasible. So we argue that transfer-learning from relevant and general domain will create a more accurate language model for EMS domain.

For general corpora, state-of-the-art BERT-Base is pre-trained using English Wikipedia and the Books Corpus (Devlin et al. 2018). The vocabulary is quite different from EMS corpora, thus rendering this pre-training corpus only is quite inappropriate. BioBERT is the first BERT model released for the biomedical domain (Lee et al. 2020) which is initialized from BERT-Base and trained using PubMed abstracts. ClinicalBERT is also a clinically oriented BERT model (Alsentzer et al. 2019) which is initialized from BioBERT v1.0 and trained with additional steps using MIMIC-III clinical notes. We use the BioBERT and ClinicalBERT vocabulary with BERT-base to augment our EMS corpora for simultaneous pre-training of EMS-BERT.

Table 5.1 summarizes the previous BERT-based dataset we use to augment our EMS corpora. Training a BERT model with a smaller corpus degrades the performance by introducing more false positives. As there is no public EMS corpus and collecting real-world EMS narrative is subject to different prohibitions, we adopted the method of simultaneous pre-training introduced by the authors in OuBioBERT (Wada et al. 2020) to increase the size and cover additional entities. Simultaneous pre-training of BERT with domain specific knowledge and generic corpora provides better results. This is achieved by increasing the frequency of pre-training instances for MLM (masked language modeling). Instead of only using the corpus from the medical domain, we use documents from the EMS domain too, with the general medical domain. Using the negative instances of NSP (next sentence prediction) where a sentence pair is constructed by pairing two random sentences from different documents, simultaneous pre-training method also increases the number of combinations of documents and enhances EMS word representations in the vocabulary.

The simultaneous pre-training technique is illustrated in Figure 5.1. This approach successfully creates an efficient pre-training corpora from multiple domains. While pre-training the EMS-BERT model, the core corpora is constituted from both the general and medical domains. The EMS corpora is considered as subordinate corpora here, which is used to create mixed training instances. During the implementation, the entire corpus was divided into smaller text files. This was particularly helpful to create simultaneous pre-training instances from different type of corpora. The combinations of NSP are determined within each split file, and the duplicate factor is set to define the number of times the sentences are used. There are two problems that arises in these cases. The first is that the duplicate factor is applied to the entire corpora of both core corpora and subordinate corpora. Thus, the smaller corpora remain relatively small. The second problem is that the combinations of NSP are limited to the file that was initially split. To solve these issues, both core corpora and subordinate EMS corpora are first divided into smaller documents with the same size for EMS-BERT. Later, we combine them to create pre-training instances. When we combined them, it was ensured that the documents in both of the corpora would be comparable in terms of their file sizes and diversity of the patterns. Using this technique, more instances from core corpora were used compared to those from subordinate corpora. With this homogeneously mixed dataset, the model achieved a higher increase in the frequency of pre-training for MLM. Using documents of core corpora for the process of pre-training creates larger training dataset than the original BERT method. It also generates an increased number of different combinations of documents compared to the original method. Core corpora and subordinate corpora were combined so that their proportion were equal, thus a higher number of pre-training instances were created to train the EMS-BERT model. Comparing with the state-of-the-art BERT models and their pre-training dataset, our dataset volume is comprehensive and provides a better accuracy for the EMS domain. Appendix C shows a sample document from out EMS corpus.

5.2.3 Fine tuning EMS-BERT

For fine-tuning, a pre-trained language model generates a set of vectors with contextual representations. A task-specific prediction layer placed on top produces the final output for the application. Task-specific model parameters are trained from the task-specific training data. While training, BERT model parameters are fine-tuned by gradient descent using back-propagation. An input instance from EMS corpora goes through task-specific pre-processing and addition of special instance markers ([CLS], [SEP], etc.). The transformed input is then tokenized using the pre-training vocabulary of the neural language model. The sequence of vectors in contextual representations taken from the language model is then processed by a feature module and input into a prediction module to produce its final output of the given task.

A sentence is transformed into an instance for BERT by replacing target entities with dummy tokens and adding special tokens. In the relation-extraction task, we use [CLS] BERT encoding as a featurizer and predict the relationship between the entities by multi-class classification. The relation extraction task predicts relations between two entities and their types mentioned in the sentence. We explored three entities from the EMS corpora - signs and symptoms, medications, and interventions. Our experiments predicted all the six pairs of relations among these three entities from the textual corpora.

Utilizing the approach discussed in the BLUE benchmark (Y. Peng, Yan, and Z. Lu 2019), this task is implemented as a sentence classification task by using anonymous entities within the sentences and predefined tags such as @SYMPTOM$ and @INTERVENTION$ (Lee et al. 2020). Figure 5.2 shows a general architecture of fine-tuning a BERT model for downstream tasks (Wada et al. 2020). We fine-tune EMS-BERT for the following two tasks: (i) EMS concept recognition, and (ii) relation extraction. Compared to regular clinical corpora, EMS concepts cover a wide range of clinical conditions, medications, and intervention. These entities may be correlated depending on the recovery protocol. Relation extraction from EMS corpora signifies such dependencies and infers missing information from the narratives. Thus, accuracy of relation extraction task highlights potentially missing attributes from a given set of EMS interventions. For comprehensiveness of the evaluation, we compare EMS-BERT with state-of-the-art clinical concept detection tools, and with relevant BERT models for clinical and medical domain. Two evaluations are detailed in the following sections. First, we studied the EMS entity/concept recognition using state-of-the-art clinical concept recognition baseline tools such as MetaMap (Aronson and Lang 2010), cTAKES (Savova et al. 2010), and CLAMP (Soysal et al. 2017).

Second, we showed the accuracy of relation extraction using EMS-BERT using the ground truth developed by EMS professionals.

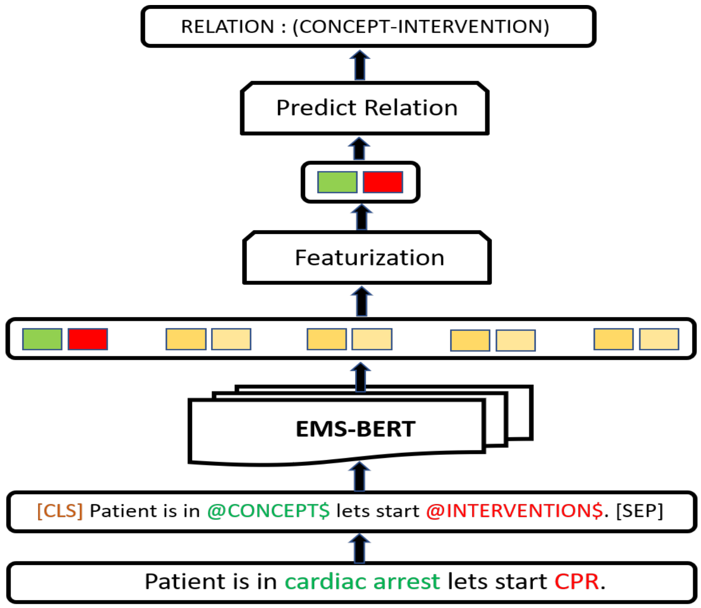


Figure 5.2: A fine-tuning example for EMS-BERT model

5.3 Evaluation of downstream tasks

In this section, we describe the experimental setup and dataset used to pre-train and test EMS-BERT. We also present the results of the experiments for recognition of EMS entities, relation extraction, and inferring missing information.

5.3.1 Experimental design and dataset

Setup

We use mixed precision training of FP16 computation for both pre-training and fine-tuning EMS-BERT. This method accelerates the computation significantly compared

Table 5.1: Dataset for EMS-BERT

Model	Corpus	Number of Words	Domain
BERT-base	English Wikipedia, Book Corpus	2B	General
BioBERT	Wikipedia, Books, PubMed, PMC	3.5B	General, Medical
ClinicalBERT	MIMIC-III (Health Records)	1.5B	Medical
EMS-BERT	EMS Corpora	2.4M	EMS

to other methods as it uses half-precision format. Two NVIDIA RTX-8000 of 32 GB size GPU are used for pre-training; a single GPU is used for fine-tuning. The configuration and weight initialization are almost same as the BERT-base. We modified the NVIDIA implementation to utilize FP16 computation, gradient accumulation, and a layer-wise adaptive based optimizer (LAMB) (You et al. 2019). For pre-training, we set the maximum sequence length of 128 tokens and trained the model for 5,068 steps using the global batch size (GBS) of 65,536 and a LAMB optimizer with the learning rate (LR) of 6e–3. Subsequently, we continued to train the model allowing a sequence length up to 512 tokens for an additional 1,272 steps to learn positional embeddings. The size of the amplified vocabulary is 30,700.

For EMS entity recognition, EMS-BERT performs sequential labelling and detects the required entities in the given text. The BERT encoding of a given sequence of token predicts the label and recognizes the entity. The relation extraction task predicts relations between two entities and their types mentioned in the sentence. We explored three entities from the EMS corpora - signs and symptoms, medication, and intervention. Our experiments predicted all six relations among these three entities from the textual corpora. We also avoid overfitting by inserting dummy tags for entities, as depicted in Figures 4.1 and 5.2. Using the relation extraction task and a mapping of prerequisites of different entities developed by certified EMS professionals, we infer potentially missing information from the EMS test set. This information depicts how thoroughly each of the approaches cover the EMS entities in an EMS document.

Dataset and metrics

Table 5.2: Entity/Concept recognition using EMS-BERT

Method/Metric	Precision	Recall	F-1 score
BERT-Base	52.58	51.21	51.89
EMS-BERT-wsn	64.87	61.09	62.91
CLAMP	69.29	62.58	63.42
cTAKES	62.94	65.81	64.34
KnowBERT	67.34	66.17	66.75
ClinicalBERT	69.59	68.24	68.91
MetaMap	71.23	68.47	69.82
BioBERT v1.1	73.84	70.81	72.29
EMSContExt	74.62	71.54	73.05
EMS-BERT	81.24	76.59	78.85

We utilized some of the datasets used in BERT-base (Devlin et al. 2018), BioBERT (Lee et al. 2020) and ClinicalBERT (K. Huang, Altosaar, and Ranganath 2019) to pre-train EMS-BERT. BERT-base use English Wikipedia and Book Corpus as general domain corpora. BioBERT and ClinicalBERT use PubMed abstracts and PubMed Central Full-Text articles (PMC) (McEntyre and Lipman 2001), and Medical Information Mart for Intensive Care III dataset (MIMIC-III) (Johnson et al. 2016), respectively. These two datasets hold information specific to the medical and clinical domain. For EMS-BERT, we create our simultaneous training instances by combining the general, medical and clinical domain information with EMS corpora (depicted in Figure 5.1). Table 5.1 summarizes the datasets used to pre-train EMS-BERT. For EMS corpora, we used 36,000 EMS narratives for creating simultaneous training instances and 4,000 annotated EMS narratives for validation and testing EMS-BERT. The testing set includes both noisy, unstructured EMS transcripts and structured, post-scene EMS narratives. Certified EMS professionals supervised the annotation of the EMS dataset. Since our target is to measure how accurately EMS-BERT recognizes EMS entities, i.e., signs and symptoms, medications, and interventions, and extract relations between each of the entity pairs, we have selected *Precision, Recall and F-1 Score* as our accuracy metrics. We also compare EMS-BERT's simultaneous pre-training method with a knowledge integration based approach known as KnowBERT (Peters, Neumann, Logan IV, et al. 2019). KnowBERT integrates knowledge bases into BERT using knowledge attention and a recontextualization component (KAR).

5.3.2 Experimental results

In this section, we detail the results obtained with EMS-BERT using the augmented dataset. We then compare these results with other state-of-the-art techniques and tools from the literature. For ablation studies of simultaneous pre-training, we also used a different pre-training of EMS-BERT which does not utilize a simultaneous pre-training method. The results obtained with this version is labelled under EMS-BERT-wsn to show the efficacy of simultaneous pre-training for our corpus. The following example helps to understand the evaluations. Let us consider the following portion of a sample narrative.

"The patient is in cardiac arrest so we start CPR compressions. I am going to do BVM bag valve mask. Check for chest rise..."

In our evaluations, **cardiac arrest** is extracted as an EMS concept of type *"signs and symptoms"*, and **CPR compressions** is extracted as of type *"intervention"*. The first relation is detected between **cardiac arrest** and **CPR compressions** in the *"signs and symptoms - intervention"* category. For missing data prediction, if the narrative did not include information regarding **chest rise** concept after **BVM bag valve mask** intervention, EMS-BERT documents a flag using the protocol dependency chart.

EMS entity recognition

Table 5.2 shows the overall average scores of EMS entity recognition, i.e., of signs and symptoms, interventions, and medication. For the noisy and structured test dataset, the average F-1 score for EMS-BERT is 78.85 (72.91 and 81.68, respectively). EMS-BERT outperforms the other state-of-the-art tools by at least 5%. Comparison with EMS-BERT-wsn emphasizes the significant of simultaneous pre-training for the EMS corpora. The average F-1 score for EMS-BERT-wsn is only 52.91. We observe that for BERT-Base, which is pre-trained on only the general domain corpus, the result if very poor. The average F1-score is 51.89 for BERT-Base which is significantly lower than that of the other state-of-the-art models. On the other hand, BioBERT v1.1 achieves higher scores than ClinicalBERT for the EMS dataset. The better results of BioBERT v1.1 is due to the higher similarity of EMS corpus with PubMed abstracts and PubMed Central Full-Text articles (PMC), compared to the Medical Information Mart for Intensive Care III dataset (MIMIC-III). For KnowBERT, the knowledge integration approach shows good results with an average F-1 score of 66.75. However, our insight suggests that the non-overlapping entities of the EMS domain and other medical, clinical domain plays a significant role for this relatively lower score. All these low scores of the other BERT models on the EMS dataset can also be attributed to the following generic reasons: (i) the lack of a silver-standard dataset for training previous state-of-the-art models, and (ii) different training/test set splits used in previous work which were unavailable. For clinical concept recognition tools

Table 5.3: Relation extraction using EMS-BERT

Relation	Metric	BERT-Base	KnowBERT	BioBERT v1.1	EMS-BERT
Signs & Symptoms - Signs & Symptoms	P	60.51	73.51	77.29	79.58
	R	58.57	70.32	72.49	73.91
	F-1	59.52	71.88	74.81	76.64
Signs & Symptoms - Intervention	P	62.54	71.61	74.38	77.92
	R	59.78	70.87	71.94	76.51
	F-1	61.13	71.21	73.14	77.21
Signs & Symptoms - Medication	P	61.75	72.84	75.87	78.62
	R	57.39	69.74	74.94	75.54
	F-1	59.49	71.26	75.40	77.05
Intervention - Intervention	P	67.31	78.61	77.31	80.67
	R	64.32	74.62	73.88	77.92
	F-1	65.78	76.56	75.56	79.27
Intervention - Medication	P	64.38	66.94	71.62	73.94
	R	65.90	65.17	69.58	71.39
	F-1	65.13	66.04	70.59	72.64
Medication - Medication	P	51.84	60.25	63.57	68.67
	R	50.68	54.39	59.84	64.31
	F-1	51.25	57.17	61.65	66.42

such as MetaMap, CLAMP, and cTAKES, these tools exhibit a high false positive rate. One possible reason is the over-generalization of entities. A semi-supervised approach such as EMSConExt shows a better F-1 score compared to these three tools, but EMS-BERT also outperforms EMSContExt.

Entity relation extraction

The relation extraction results of different BERT models are shown in Table 5.3. We predict the relations between the following three EMS entities - signs and symptoms (S.&S.), intervention (Int.) and medication (Med.). EMS-BERT achieved better performance than the other state-of-the-art models. On average, EMS-BERT obtained a higher F1 score (2%-5% higher) than original BERT-Base, KnowBERT and BioBERT v1.1 on EMS dataset. Table 5.3 shows that the accuracy of relation extraction for medication - medication is lower compared to the other pairs of relation. This is due to the inability of the model to detect other connecting contexts of the medications among themselves. Certain medications are labeled as prohibited with each other, however for the other pairs it is difficult to create a connection unless all of the underlying contexts are explicitly mentioned in the corpus and detected by the model. So any false positive or false negative in all other relation extraction may affect the relation extraction of multiple medications. For the other pairs of entities, the relation is much more straight forward, and EMS-BERT shows better results for those. Other models such as BioBERT v1.1 were not pre-trained using the augmented EMS corpora, so the lower performances are expected. KnowBERT shows good results compared to BERT-Base, this is because of the knowledge integration from the augmented EMS dataset. But compared to the results of EMS-BERT, the F-1 scores for relation extraction for each pair of entity are 5-12% lower on average. For BERT-Base, pre-training only with general domain corpus negatively impacts the outcome of relation extraction from the EMS corpus. The F-1 scores for relation extraction are 9-20% lower for each entity pair.

Inferring missing concept/entity

Table 5.4: Total coverage of related EMS entities/concepts

Approach/Metric	Coverage (%)
ClinicalBERT	77.96%
BioBERT v1.1	84.47%
EMS-BERT	91.21%

To measure the detection of missing information, we calculate the total coverage of concepts by EMS-BERT and other methods. Then, a concept dependency model developed by certified EMS personnel predicts the missing information by comparing with the output of all the methods. For each of the entities found in an EMS narrative, there are some other entities which are correlated and expected to be preceded/followed in an EMS narrative. These are often prerequisites and post-requisites of various interventions and medications. Sometimes, they are not mentioned in the transcript or post-scene narrative. Using each of the entities in our test set, our EMT collaborators developed a document with dependencies among the EMS entities. When an entity is detected by EMS-BERT, it checks the list of all the correlated entities against the detected entity and infers the potentially missing entities in the original transcript. For example, a cardiac arrest protocol which exhibits the intervention CPR, must also have information regarding an IV intervention in the corpora. Table 5.4 shows the comparison of BERT, BioBERT v1.1 and EMS-BERT for recognition of all possible EMS entities which are correlated. Here, EMS-BERT shows highest accuracy for inferring potentially missing information by detecting the maximum number of entities and their potentially correlated missing entities correctly. EMS-BERT outperforms the other two models by at least 7% for overall data dependency capture. This improvement is also significant for understanding the context of the situation and providing personalized patient care in latter stages of recovery. Inferring potentially missing information from live EMS transcripts and post-scene narratives lead to better EMS training and performance.

5.4 Discussion

For the data mixing strategy and ablations, we do not have any ablation study at the moment to support the equal nature in core and subordinate corpora for augmenting the dataset. We have the data mixing research as a future goal for the project. Our future study will target finding what proportion of mixing both kinds of corpora yield best results, and whether there exist other approaches for data augmentation with similar or better results. Different methods for augmenting a dataset with amplified vocabulary exist in the literature, such as LSTM and transfer learning based approaches. In this study, we adopt the simultaneous pre-training method and compared it with multiple knowledge base integration by KAR methods (Peters, Neumann, Logan IV, et al. 2019). As a future milestone of this research, we will investigate other data augmentation methods, run more comprehensive ablation studies for simultaneous pre-training, and compare their results with our current approach. Wolf et al. in (Y. Wang et al. 2020) discussed the construction of the uncased vocabulary via byte-pair encoding (BPE) (Sennrich, Haddow, and Birch 2015) using tokenizers. We implemented the uncased vocabulary as a custom vocabulary to suit a small corpus. A small corpus often shows biases towards subordinate corpora. To solve this problem, we amplified the core corpora and made the corpus size the same as that of the subordinate corpora. The authors in (H. Wang et al. 2021) presented Bidirectional LSTM and BERT approaches to detect entity from EMS audits from Singapore Civil Defense Force. However, our EMS dataset is comparatively unstructured, noisy and a portion of it is created from live transcripts from real EMS scenes. The authors in (H. Wang et al. 2021) mentioned that one probable reason for their low scores with BiLSTM is the inability to handle misspelling in the dataset. Our hypothesis for developing EMS-BERT precisely highlights this condition of our dataset. The authors used basic BERT-Base and ClinicalBERT models instead of developing a custom BERT model. As we are focused to develop a generic model to detect EMS concepts and understand their correlations, we concentrated on developing a custom BERT for EMS domain. Our future goal also includes using EMS-BERT for other downstreaming tasks such as negation detection, vitals validation, etc. from EMS corpora.

In this chapter, we present a study for augmenting the dataset compared to using only the EMS corpora without data augmentation. The results of EMS-BERT without simultaneous pre-training and data augmentation are documented and the experiment results show significant improvement when the simultaneous pre-training method is applied. We show that EMS-BERT outperforms ClinicalBERT and BioBERT for entity recognition, relation extraction, and inferring missing information for our EMS corpora. However, ClinicalBERT and BioBERT were developed for the medical and bio-medical domain. They are not pre-trained for the EMS domain. A more comprehensive comparison with a BERT model specifically pre-trained on EMS corpus will further strengthen the significance of EMS-BERT and it's simultaneous pre-training technique. For application of EMS-BERT, there are multiple potential scope. Cognitive assistants developed for the emergency response domain may leverage from deploying EMS-BERT in their backend. For example, different types of cognitive assistants for the emergency domain such as (S. Preum et al. 2019a; Sarah Masud Preum, Shu, Ting, et al. 2018), automated form filling (Rahman, Sarah M Preum, et al. 2020) and other applications require clinical and medical entity detection from an EMS corpus. EMS-BERT can be very effective for such assistants and applications. These previous systems used different clinical concept detection tools, but our experiments clearly indicate better F-1 scores with EMS-BERT for the EMS domain. Currently, we are working on adapting the EMS-BERT model for real-time applications. We envision developing a cognitive application for introducing automation in EMS training. EMS-BERT will be deployed to detect different concepts in this application. The application will provide customized suggestions and feedback according to the severity level of the training and experience level of the first responder.

5.5 Conclusion

To the best of our knowledge, EMS-BERT is the first language model specialized for the EMS domain. For amplifying the existing EMS corpus which consists of post-scene EMS narratives and live-transcripts, EMS-BERT also utilizes general, clinical, and medical corpus from state-of-the-art BERT, BioBERT and ClinicalBERT models. Using simultaneous pre-training technique on the amplified vocabulary, we demonstrated that a practical BERT based model can be constructed for EMS downstream tasks. Our thorough experimentation also demonstrates that EMS-BERT outperforms the existing state-of-the-art medical and clinical models by at least 2% to as much as 11% for F-1 scores in downstream tasks such as entity recognition, relation extraction, and inferring missing information on EMS domain. Even though there is room for improvement for the accuracy, the results suggest that EMS-BERT can successfully handle the complex challenges, i.e., unstructured, sparse, noisy, and high-dimensional dataset for text-mining related tasks in EMS domain. EMS-BERT also emphasizes the significance of a specialized BERT based language model for EMS specific corpus, and distinguishes the EMS domain from medical and clinical datasets.

Chapter 6

SenseEMS & EgoCap

Wearable computing devices such as smartwatches and smart-glasses are becoming popular now-a-days. The EMS domain can benefit by using sensor and vision based datasets from these gadgets. EMS providers wear smartwatches while using their hands extensively for the rescue operation and providing care to the patients in an EMS scene. To utilize computer vision techniques on smart-glass based images or video data for scene understanding from the EMS responders' view point, exploring first-person captioning on images is important. An egocentric of first-person captioning image dataset provides machine vision of the notion of "self". In this chapter, we present two models, SenseEMS and EgoCap, which address using smartwatch based sensor data for providers' hand operated EMS intervention gesture detection and real-time monitoring, and an image dataset with ego-captioning with fusion of contextual cues which exhibits situation-aware captioning, respectively. SenseEMS uses a hybrid deep neural network with appropriate real-time algorithms using the accelerometer, gyroscope, and magnetometer data to detect hand operated activity gestures, i.e. CPR compressions, and to provide real-time quality assessment on different metrics of the activity, i.e., the rate of CPR compressions. Our results for this ongoing research show promising accuracy. SenseEMS currently detects CPR rate with less than 4% error and a F-1 score of CPR compression related gesture detection using the hybrid deep neural network is 90For EgoCap, this research creates 2.1K ego-images, over 10K ego-captions, and 6.3K contextual labels, to close the gap of lacking ego-captioning datasets. This dataset is unique compared to the state-of-the-art, as EgoCap incorporates contextual labelling with first-person captioning. The dataset is diverse, which makes it a comprehensive candidate for developing models from an egocentric perspective with emphasis on the responder's activity and position.

6.1 Problem, Challenges and Overview

Wearable devices, such as smartwatches and smart-glasses, offer specific advantages for emergency medical services (EMS). Smartwatches equipped with various sensors can continuously monitor hand movements of the care providers. This data can be transmitted in real-time, allowing the providers to assess the quality of their ongoing hand operated activity and make informed decisions about the required medical interventions. Smart-glasses provide another hands-free way for EMS personnel to communicate with colleagues or medical professionals during emergencies. Responders can use smart-glasses to record the scene images and transmit live video feeds, allowing remote experts to provide guidance and support. Using computer vision techniques, the access to relevant image data from an EMS scene can assist in making situation-aware and informed decisions for providing appropriate care. In this chapter, we present the analysis and results of two of our ongoing research projects, SenseEMS and EgoCap. SenseEMS is developed for EMS providers' hand activity gesture detection and parameter monitoring using smartwatch based sensor data. EgoCap is an image dataset with egocentric captioning to bridge the gap between state-of-the-art visual EMS assistants and automated egocentric contextual scene narration from live images for emergency responders.

6.1.1 SenseEMS - using smartwatch based sensor data for hand activity gesture detection and parameter monitoring in EMS

From the perspective of smartwatch based sensor data, EMS providers use their hands extensively to complete various interventions during an EMS rescue. Hand operated interventions include attaching different equipment on the patient's body, administering medications, and performing life-saving procedures such as CPR compressions. Most of these hand operated interventions contain dynamic parameters from both the provider's and patient's point of view. For example, CPR compressions and its parameters vary according to the age of the patient. For CPR, the first responder needs to be switched after a certain period of time to ensure proper quality of the compression. An EMS provider has to go through different levels of training before performing these interventions in a real scene. Case studies in the U.S. show that the EMS programs rarely use any assistive technology for quality assessment. Even today, the training for hand operated interventions are guided manually (Hobbs 2020). To improve the quality of interventions and to classify sensitive hand operated emergency interventions, real-time and automated technologies can be adopted for EMS training sessions and actual scenes (Sarah Masud Preum, Munir, et al. 2021). To this end, we present in this chapter our ongoing research **SenseEMS**, an automated assistant which uses smartwatch based sensor data for detecting and monitoring hand operated EMS gestures for interventions such as CPR compressions. This is a challenging problem because different responders move their hands differently. There are many confounding hand moving gestures, and even the activities of interest have many similar motions to each other for different types of patients.

Due to the popularity of smartwatches and necessity of tracking time, first responders wear smartwatches on their wrist during EMS training and real scenes. This enables us to easily collect acceleromoter, gyroscope, and magnetometer data when they are performing EMS interventions. Leveraging state-of-the-art deep learning networks and appropriate real-time algorithms, we can process these sensor data for automated classification of different EMS intervention gestures and activities, and assessment of quality for specific parameters of the hand operated activity. Previous research (Wen, Ramos Rojas, and A. K. Dey 2016; Samyoun et al. 2021) have addressed finger and hand motion detection using sensor data from different sources. However, the EMS domain remains unexplored for the usability of smartwatch based sensor data for activity detection and monitoring. Our goal is to separate the interventions of interest from regular hand movements and provide automated quality assessment with real-time parameter prediction. This greatly benefits the EMS training procedure, and improves the real-scene application. Figure 6.1 shows an overview for the problem and our proposed solution.

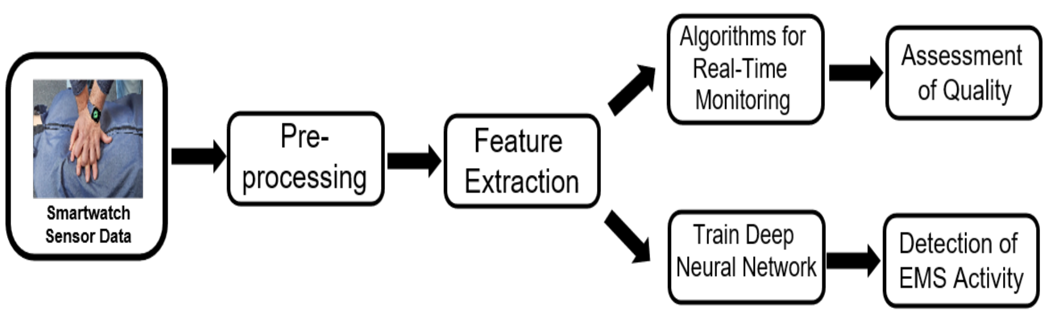


Figure 6.1: Overview of using smartwatch based sensor data for hand operated activity detection and monitoring in EMS

6.1.2 EgoCap - an image dataset with egocentric captioning

Applying computer vision techniques in EMS domain has become increasing feasible as vision data collected by body-worn cameras have seen a dramatic surge in the past decade. These image data contain valuable information about the camera person's status as well as the surroundings. Although object-detection oriented scene understanding has accomplished tremendous success, egocentric vision data are typically contaminated by motion blurring, hand occlusion, and awkward camera angles (al. 2021). Describing egocentric vision data using natural language, also called ego-captioning, is currently an active research topic. Ego-captioning aims at human-understandable interpretation of vision data which is crucial for various life-logging applications such as EMS. Prospective use cases include auto calorie intake recording for people on a diet (Bolaños, Dimiccoli, and Radeva 2017), daily activity tracking for patients (Fan and Crandall 2016), and event summarization for emergency responders (Rahman, Sarah M Preum, et al. 2020). As shown in Figure 6.2, first-person captioning provides a precise perspective in storytelling, whereas, a third-person narrative poses ambiguity. Moreover, a first-person narrator places the viewer at the centre of the action and lends credence to the narration. The first-person perspective establishes rapport with readers by sharing a personal narrative with them directly. Thus, ego-captioning is also critical for artificial intelligence to establish the notion of "self".

The state-of-the-art data-driven captioning has largely focused on describing the contents objectively, such as third-person narrative (W. Liu et al. 2021; Anderson et al. 2018). This results in most captioning datasets being labelled in the third person, such as COCO (Lin et al. 2014) and MSR-VTT (Xu et al. 2016). In this chapter, we present our ongoing research to ego-image captioning techniques and how we plan to

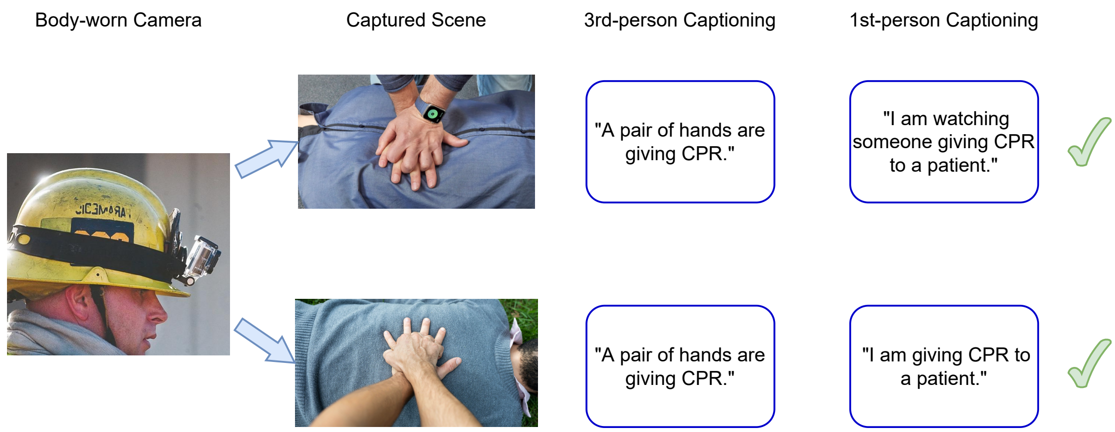


Figure 6.2: In EMS, first-person captioning will resolve ambiguity where third-person fails.

use the model for EMS domain. First-person captions cannot be easily created from third-person captions, as syntax of a first-person narrative is semantically different from a third-person narrative. Empirical implications of the camera person's status have to be made as the cameraperson is usually outside the field of view. To this end, here we present a new dataset, **EgoCap**, comprising life-logging images with five ego-captions each to generate first-person captions consistently. We select source images from prevailing datasets - COCO (Lin et al. 2014), MSVD (D. L. Chen and Dolan 2011), MSR-VTT (Xu et al. 2016), and Ego4D (al. 2021)) to avoid privacy issues and to increase scene diversity. EgoCap incorporates contextual labels such as *where*, *when*, and *whom*, through querying surveyors. Due to privacy reasons, and the time-consuming nature of creating a dataset from scratch, we use EMS relevant images from these public repositories instead of collecting images from actual EMS scenes.

6.2 Approach and Solution

In the following subsections, we detail the current solution of SenseEMS and approach for creating the EgoCap dataset.

6.2.1 Methodology for SenseEMS

To facilitate automated and improved learning experiences during the training, and better performances in real-EMS scenes for the EMS providers, SenseEMS uses smartwatch based sensor data for the detection and quality assessment of CPR compressions. This thesis only presents the details on gesture detection for CPR, and real-time rate estimation for assessment of CPR quality. For this research, we use Samsung Galaxy Smartwatch5 and Asus Zenwatch2 models for collecting data from accelerometer, gyroscope, and magnetometer sensors. We use an android app WaDa(Mondol et al. 2018) to collect the sensor data. The sensor readings are collected with timestamps throughout the event at 50Hz sampling rate. Before processing, several statistical features are extracted from the data. Each of the sensors provides data signals along the X, Y, and Z axes. Pre-processing is required to remove the noisy artifacts from the sensor readings. Specifically, we pass the raw signals through a finite impulse response filter to remove the high-frequency vibration noises. Window size of 0.1s and an overlap of 50% is selected for the training purpose. Statistical features are generated from each window, i.e., the mean, standard deviation, kurtosis, and skew feature. For classifying CPR compressions gestures, the features are fed to a hybrid learning model which is a parallel combination of CNN and RNN (Samyoun et al. 2021). Figure 6.3 shows the high level architecture of the classifier model. We use different number of filters for convolutional layers. The parameters of the model are chosen based on a preliminary evaluation on a validation set of our overall sensor dataset. Combination of CNN and RNN allows capturing the spatial and temporal correlation present in input sensor data. As a result, the combination of these two networks identifies the continuous activity gestures with high accuracy. For detecting the CPR rate, SenseEMS uses dispersion based peak detection using Z-score on the standard deviation feature. CPR rate is calculated from the average time between consecutive peaks.

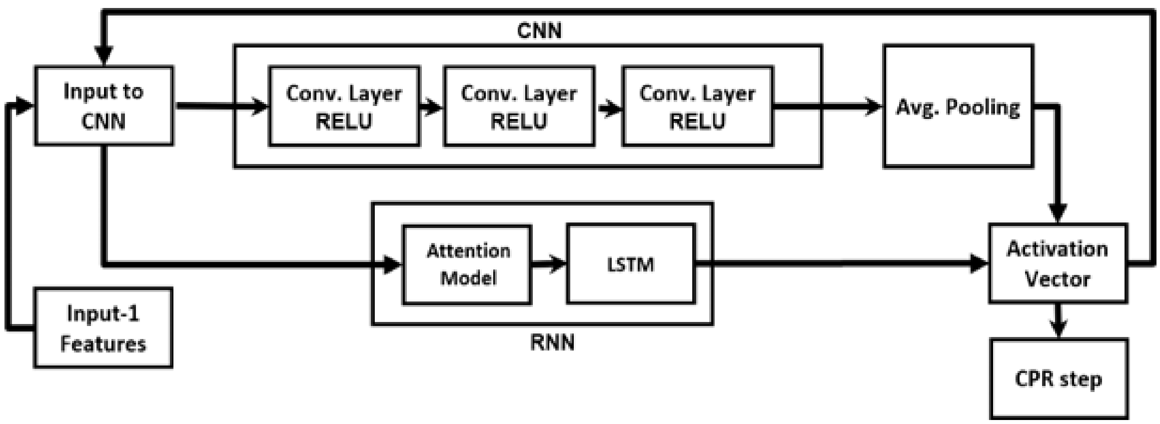


Figure 6.3: A hybrid attention model with deep neural network for EMS activity (CPR) detection

For CPR, a complete compression cycle consists of following two gestures- (i) compression on chest via downward movement of palm(s) for specific time and depth, and (ii) upward rebound for specific time and height. These two gestures are depicted by ① and ② signs in the Figure 6.4, along with different types of compressions for different patients. Table 6.1 shows details of parameter variation for CPR.

6.2.2 Methodology for Creating EgoCap Dataset

To create the first-person captioning for EgoCap image dataset, each image is given five captions in only egocentric narrative alongside three contextual labels - *where*,

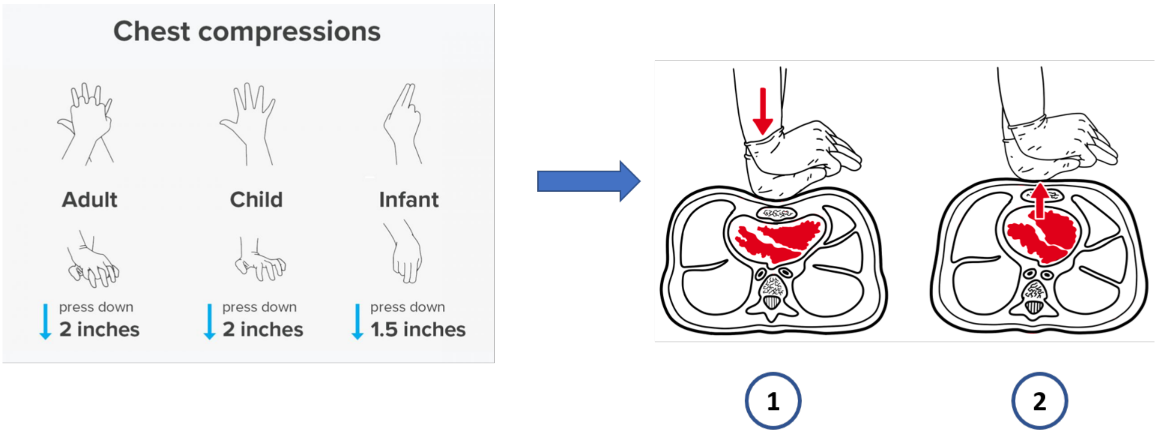
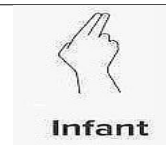


Figure 6.4: Basic types and gestures for CPR

when, and *whom*. EgoCap consists of 2.1K images, covering most daily scenes. Due to the unavailability of any large and public EMS dataset, we utilize images or video frames from widely acknowledged datasets as sources to maximize scene diversity. These source images with EMS scene relevance are collected from datasets widely acknowledged in visual-semantic studies as well, including COCO (Lin et al. 2014), MSVD (D. L. Chen and Dolan 2011), MSR-VTT (Xu et al. 2016), and Ego4D (al. 2021). This not only maximizes scene diversity, but evades privacy concerns for release, which would have been an issue if we attempted to work with EMS based images. We shuffle the sources and handpick images that strictly conform to the shooting angle of a wearable camera, and incorporate multiple categories defined in original sources. The labeling was conducted by five surveyors with expert knowledge in computer vision, NLP, and visual captioning. The annotations are further verified through a review and correction process. The context labels are regarded as probability distributions for contextual representation learning.

Table 6.1: Types and parameters of CPR.

Patient Type	**Depth (inches)**	**Count in Each Cycle**	**Rate**	**Compression Ventilation Ratio**	**Reference Gesture**
Adult ($\geq$ 12 years)	2	30	100-120/min	30:2	
Child (1-12 years)	2	30	100/min	30:2	
Infant ($\leq$ 1 years)	1.5	30	100/min	30:2	

Source Image Selection

One of the reasons for choosing images from above mentioned sources is that these sources come with either third-person captions or Human Activity Classification (HAC) labels which can be used for reference. We take into account these soft references to optimize image theme distribution and HAC distribution as shown in Figure 6.5. It can be seen that more than one third of the images capture interaction with people or salient object(s) in sight, while the rest are casual shootings without clear themes. We also define a set of criteria to exclude images that are not deemed ego-images, i.e., staged scenes or cartoons. No scripted scenes are allowed, and blurred or hand-occluded frames are incorporated without bias to reflect real-world noises embedded in ego-images. A reporting mechanism is implemented to allow surveyors to skip the image if it is found inappropriate for ego-captioning.

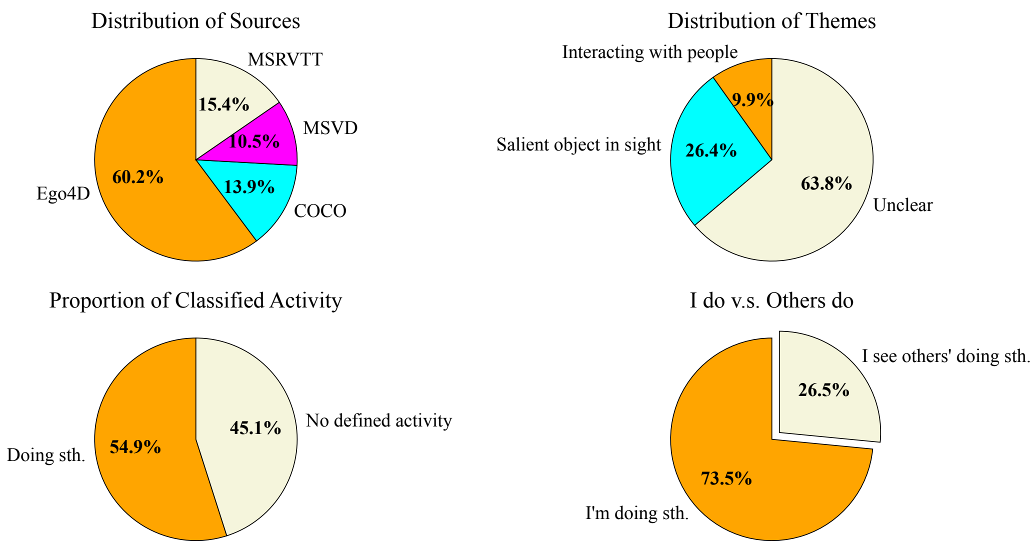


Figure 6.5: The composition of EgoCap (top left), Distribution of image themes (top right), Proportion of images with explicitly labelled activity or not (bottom left), and Proportion of “I do” versus “I see others do” in those with classified activity labels (bottom right).

Annotations

User Interface. The core idea of EgoCap dataset is to label the images using qualified and diverse first-person narratives. Similar to COCO, we aim to create five captions for each image. In doing so, we assign five surveyors, who are trained to write ego-captioning following specific guidance. To minimize repetitions, the surveyors cannot see each other’s captions. The source images are designed to emerge in random orders to reduce the fatigue factor. In order to streamline and standardize the user input, we create a web-based graphic user interface (GUI) to allow easy logging and formatting of the captions. An example of the annotation panel is shown in Figure 6.6. We set explicit instructions and visual aids to guide the surveyors in captioning. In short, we expect the surveyor to consider as the cameraperson and to use a sentence to describe own status, activity, current location, etc. The caption

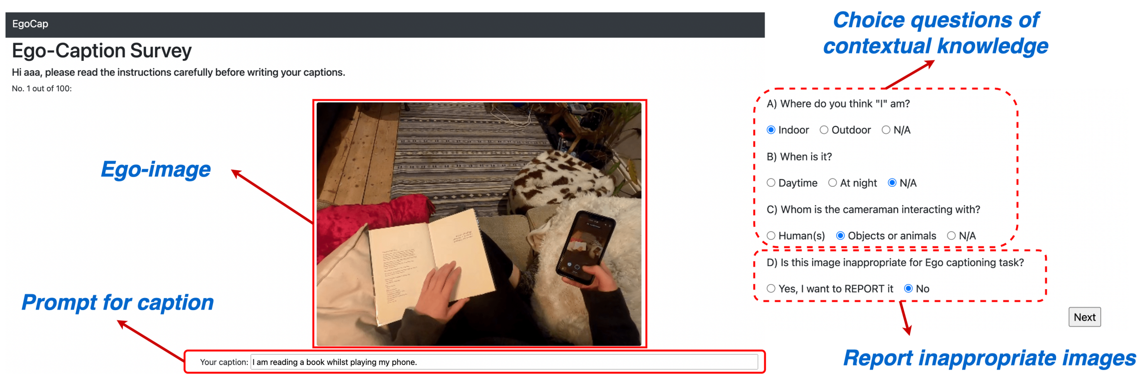


Figure 6.6: An illustration of the EgoCap user interface prompting inputs of ego-captions and contextual knowledge.

should focus on "my" activity, "my" interaction with the object or creature in sight, and "my" surroundings. The labeling task was divided into 21 splits each of which comes with 100 ego-images.

Contextual Information. We also poll contextual information for each image. These tagging options include (a) *where*: am "I" indoor, outdoor, or ambiguous; (b) *when*: is it daytime, at night, or ambiguous; (c) *whom*: am "I" interacting with human, object/animals, or ambiguous. This is designed to collect auxiliary contextual facts about the ego-images from human intelligence.

Label Verification. It is common that the surveyors misunderstand the scene or some samples draw disagreement. We adopt an anomaly filtering and reviewing process to remove inappropriate samples and correct captions. There are two mechanisms designed for this. Firstly, the labels are sorted by mutual similarity measures to reveal potential irrelevant captions. This is performed on the basis of taking turns to use one candidate caption as hypothesis and the others as references to compute average BLEU (Papineni et al. 2002) scores. The inconsistent captions were updated. Secondly, a reporting button allows a surveyor to mark as irrelevant if a source image is found inappropriate, i.e., a synthesis cartoon frame. These potentially irrelevant samples were picked out and reviewed on qualification.

6.3 Results and Analysis

In this section, we compare SenseEMS with two other existing methods for hand gesture detection and present our preliminary findings for CPR rate detection using z-score method on standard deviation of sensor data. We further analyze EgoCap by comparing the dataset with other state-of-the-art dataset as well.

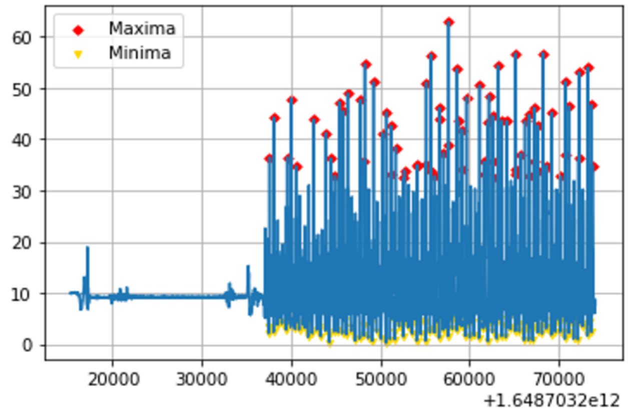
6.3.1 SenseEMS Comparisons

We collected smartwatch based sensor data from 20 EMT participants and 20 non-EMT participants with previous experience of performing CPR. The participants wore the smartwatch on their dominant wrist while performing all the three different variants of CPR. The participants were aged from 24-50 years old. Half of the participants were male and the other half were female. Each participant provided data for 5 minutes to evaluate the accuracy of CPR gesture detection and rate estimation model.

Table 6.2: Different methods for CPR hand gesture detection

Metric/Tools	bi-LSTM (Zhu et al. 2018)	SVM (Wen, Ramos Rojas, and A. K. Dey 2016)	SenseEMS
Precision	0.88	0.84	0.92
Recall	0.87	0.79	0.89
F-1	0.87	0.81	0.90

Figure 6.7 shows our results for CPR rate estimation using accelerometer data and



Ground Truth Rate	Detected Rate
121.5	122.77

Figure 6.7: CPR rate estimation based on Accelerometer data

peak detection algorithm. Table 6.2 shows accuracy of SenseEMS for CPR compressions gesture detection against other state-of-the-art techniques. SenseEMS records highest average F-1 score of 90% for our current test dataset, compared to other solutions based on bidirectional LSTM (Zhu et al. 2018) and SVM (Wen, Ramos Rojas, and A. K. Dey 2016) based networks. This result only highlights the basic gesture detection for CPR. In future, we aim to combine the gestures for CPR and other hand operated activity classification from continuous sensor stream.

6.3.2 EgoCap Dataset Analysis

After reviewing and anomaly removal, we selected 2079 verified images for captioning (1252 from Ego4D; 289 from COCO; 218 from MSVD; 320 from MSR-VTT). Finally, EgoCap comprises over 10K ego-captions alongside 6.3K contextual tags. We also retrieve weak labels of third-person captions or HACs from their source datasets, and associate them for reference. To the best of our knowledge, this is a first sizable dataset, with labelled contextual information, that allows end-to-end ego-caption learning. We compare EgoCap to existing datasets for size and different features

Table 6.3: A comparison of existing egocentric or captioning datasets.

Datasets	Size	Labels					
		Diverse	*OD*$^\diamond$	*HAC**	Third-caption	First-caption	Context
*COCO*Lin et al. 2014	$118K$	✓	✓		✓		
*MSVD*D. L. Chen and Dolan 2011	$1.9K$	✓			✓		
*MSR-VTT*Xu et al. 2016	$10K$	✓			✓		
*Charades-Ego*Sigurdsson et al. 2018	$4K$			✓			
*EPIC-Kitchens*Damen et al. 2018	$100\,h$		✓	✓			
*Deepdiary*Fan, Zhang, and Crandall 2018	$7.7K^\triangle$			✓		✓	
*EDUB-SegDesc*Bolanos et al. 2017	$1.3K^\dagger$					✓	
*Ego4D*Dal. 2021	$3025\,h$	✓		✓			
EgoCap	$2.1K$	✓				✓	✓

$\diamond$ Object Detection.

$\star$ Human Activity Classification.

$\triangle$ Fewer than 300 images are released for privacy concerns.

$\dagger$ Unavailable for download.

of labelling. As shown in Table 6.3, the size of EgoCap is currently minimal, but EgoCap is the only dataset with contextual labelling for first-person captioning. The dataset is diverse, which makes it a comprehensive candidate for developing models in safety-critical applications such as emergency response.

6.4 Conclusion

This chapter presents two challenges and solutions for sensor and visual data based assistance to EMS providers. First, we present SenseEMS, a smartwatch retrieved sensor data based method for hand activity gesture detection and monitoring in Emergency Medical Systems (EMS). Using an attention based hybrid deep neural network for detecting patterns in CPR compressions, and a suitable real-time algorithm on the extracted features, SenseEMS provides post-scene classification and real-time quality assessment for the rate of CPR during the training and real-scenes. SenseEMS currently detects CPR rate with less than 4% error and a F-1 score of CPR compression related gesture detection using the hybrid deep neural network is 90%. Surveying 31 anonymous EMS providers on SenseEMS reveals a potential influence of this research in EMS domain. Second, we create a dataset, EgoCap, with egocentric image-caption pairs with context knowledge of location, time, and saliency of the scene. EgoCap comprises over 10K ego-captions alongside 6.3K contextual tags. This is a first sizable dataset with labelled contextual information that allows end-to-end ego-caption learning. An important future work includes creating an EMS dataset for first-person captioning, and developing a situation-aware model for visual application in EMS scenes.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

Technology is revolutionizing the way emergency medical services (EMS) respond to critical situations. The use of advanced medical equipment and software in EMS has dramatically improved patient outcomes and reduced response times. For instance, GPS tracking devices, advanced communication systems, and telemedicine technologies enable EMS teams to quickly locate and respond to emergencies. Additionally, the use of advanced medical devices such as defibrillators, ventilators, and monitoring systems have greatly improved the quality of care provided to patients. Furthermore, electronic medical records have made it easier to access patient information, making it easier for EMS teams to provide the right treatment in a timely manner. Although technology plays a vital role in emergency medical services and improving patient outcomes, most of the EMS oriented training and tasks still remain under-explored from an intelligent automation point of view. To this end, this thesis attempts to invent new and innovative solutions for developing an automated and intelligent cognitive support system for the EMS care providers using computer science oriented research methods.The vision of the thesis revolves around the idea that by utilizing natural language processing (NLP) and transformer based language models on on-scene conversational audio data from Emergency Medical Services (EMS) providers and textual EMS corpus; and by effectively combining transformer-based attention techniques with deep learning and egocentric NLP captioning on data from different sensors and image data, respectively; this thesis shows how to build intelligent, interactive components of a cognitive assistant for emergency care providers, and thereby moving the state-of-the-art toward more comprehensive and automated EMS training, and for on-scene and post-scene solutions for first-responders.

Currently, many EMS tasks are manually operated, and very few researchers have addressed the challenges for developing assistive tools for automation of these tasks. For example, firstly, creating a manual summary report in EMS can be tedious and time-consuming. During emergencies, EMS teams must prioritize patient care and respond quickly, leaving little time for administrative tasks. Manual report creation can be particularly challenging, given the high-stress and fast-paced environment of EMS. Additionally, the accuracy of the report heavily relies on the memory and attention to detail of the reporting personnel. Human error is always a risk, especially when working under pressure which can lead to inaccurate information being recorded. Automated reporting systems and technology can reduce the burden of manual reporting and improve the accuracy and speed of report creation, allowing EMS teams to focus on providing critical care to patients in need. After extensive interviews with emergency responders, one chapter of this thesis has addressed the technical challenges for developing a cognitive assistant for automatically generating post-scene patient summary report using the audio data from EMS responders' communications. This component, GRACE, has relieved the care providers from cognitive overload of using their memory during and after stressed situations for the documentation purposes. GRACE is the first natural language processing (NLP) based system to address formal documentation or reporting of critical information for emergency response. Through a thorough evaluation using real EMS dataset and noise-simulated cases that includes both textual and speech EMS data, this research proves the efficacy of GRACE. The solution achieves an F1 score as high as 94%, 78%, 96%, and 83% when the data is noise-free audio, noisy audio, noise-free textual narratives, and noisy textual narratives, respectively for automated post-EMS patient summary report generation using a standard report form.

Secondly, another important aspect of EMS is training the care providers which is also currently a manually guided event. Automated assistants can play a crucial role in emergency medical responder training by providing an immersive and interactive learning experience. In a high-stress, real-life emergency situation, responders need to be able to think instantly and make split-second decisions. An automated assistant can provide personalized feedback and support to each trainee, addressing knowledge gaps and improving overall performance. Moreover, an automated assistant can significantly reduce the costs and time associated with traditional training methods, such as instructor-led training, classroom lectures, and hands-on simulations. Overall, an automated assistant can help improve the effectiveness of emergency medical responder training, ensuring that responders are better equipped to handle emergencies and save lives. For addressing this issue and associated challenges, another chapter of this thesis presents an assistant for interactive training and mock real scenes. The most frequently occurring emergency event in USA is cardiac arrest, and the research scope is narrowed down to this specific emergency event for building a comprehensive responder-customized intelligent component. This system, 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. A thorough evaluation of emsReACT 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. Extensive survey with 31 anonymous EMS providers reveal that emsReACT can play an important role in reducing the real-time cognitive overload, and creating a geographically common and scalable platform for training the EMS care providers with same standards.

The third challenge addressed in this thesis lies in detecting EMS related entities from EMS textual corpus. An EMS domain specific language model is necessary because it can provide specialized support for detecting relevant information and relation from the communication narratives of the EMS field. EMS personnel need to communicate quickly and accurately, as they often deal with life-threatening situations where every second counts. A dedicated language model that is trained specifically on EMS terminology, procedures, and protocols can greatly assist in locating and interpreting EMS-specific ontology which is different from existing medical and clinical vocabulary. Additionally, a domain-specific language model can help reduce the likelihood of misunderstandings information, and detecting potentially missing information. Ultimately, an EMS-specific language model can improve the efficiency and safety of emergency medical services, potentially discovering life saving information. For addressing these issues in EMS text processing, one chapter of this thesis presents EMS-BERT - the first language model specialized for the EMS domain. For amplifying the existing EMS corpus which consists of post-scene EMS narratives and live-transcripts, EMS-BERT also utilizes general, clinical, and medical corpus from state-of-the-art BERT (Devlin et al. 2018), BioBERT (Lee et al. 2020) and ClinicalBERT (K. Huang, Altosaar, and Ranganath 2019) models. Using simultaneous pre-training technique on the amplified vocabulary, the thesis demonstrated that a practical BERT based model can be constructed for EMS downstream tasks. EMS-BERT thorough experimentation also demonstrates that the model outperforms the existing state-of-the-art medical and clinical models by at least 2% to as much as 11% for F-1 scores in downstream tasks such as entity recognition, relation extraction, and inferring missing information using static protocol guidelines for the EMS domain.

Another chapter presents two challenges for sensor and visual data based assistance to EMS providers. These two ongoing research projects address hand activity detection and monitoring in EMS using smartwatch-based sensor data, and processing image data for understanding an EMS scene using first-person captioning of images, in particular, respectively. For different activities, smartwatches have become increasingly popular due to their ability to track various physical activities, including hand movements. In EMS, the ability to accurately recognize and monitor providers' hand activity is crucial. Smartwatch-based sensor data can provide important information about the providers' hand movements, which can aid in identifying the type of ongoing activity and guiding medical intervention. This data can also help in monitoring the type and progress of the intervention by the provider during the treatment and rehabilitation process. Moreover, smartwatch-based sensor data can enable the EMS team to track their own movements and activities during training sessions, ensuring that they are not overexerting themselves and risking injury. Therefore, the use of smartwatch-based sensor data for providers' hand activity recognition and monitoring in EMS can greatly improve provider performance outcome and ensure the safety of EMS personnel in training sessions. In this thesis, we also present our ongoing research model, SenseEMS, for hand activity gesture detection and a real-time monitoring approach using an attention based hybrid deep neural network, and dispersion based z-score calculation on features extracted from smartwatch-based acceleromoter, gyroscope, and magnetometer data, respectively. Compared to other solutions based on bidirectional LSTM and SVM networks, SenseEMS currently records higher average F-1 score of 90% on our test dataset for gesture detection on one of the most crucial and life-saving EMS intervention - CPR compression. This ongoing research also reveals less than 4% error for detecting CPR rate with SenseEMS using real-time peak detetction method. Surveying multiple anonymous EMS providers on SenseEMS reveals a potential influence of this research in EMS domain.

The thesis also discusses developing a dataset for processing image data to understand an EMS scene. This last of the thesis illustrates first-person captioning of images in particular. First-person captioning of images is important for EMS because it can provide critical information about a patient's condition that may not be immediately apparent to care providers. By capturing images from the perspective of the person providing care, first-person captioning can offer a more accurate and detailed account of the patient's injuries, symptoms, and vital signs, as well as hazardous condition of the surrounding. This can be especially important in emergency situations where time is of the essence and quick decision-making is required. Additionally, first-person captioning in real-time can serve as a valuable tool for communication between medical professionals, allowing them to share information and collaborate more effectively in order to provide the best possible care for the patient. However, there is no public EMS image dataset and collecting EMS image data requires multiple approvals and time-consuming efforts. In this thesis, first-person captioning idea has been experimented on real-life and publicly available image dataset. We create a dataset, EgoCap, with egocentric image-caption pairs and context knowledge of location, time, and saliency of potentially relevant indoor and outdoor scenes. EgoCap consists of 2.1K images with five egocentric captions each, covering most daily scenes which are common for EMS cases. The labeling was conducted by five surveyors with expert knowledge in EMS, computer vision, NLP, and visual captioning.

7.2 Future Work

We envision further research on the problems discussed in this thesis. One of the main ideas for future work includes an extensive survey with EMS responders. For GRACE, we are yet to test our system in real-world EMS scenarios. In the future we plan to highlight missing interventions and critical inconsistencies detected from the conversation regarding patient's clinical condition. We also aim to develop a more generic and scalable approach by considering multi-patient and multi-responder scenes, and by applying suitable machine learning techniques.

For emsReACT, the 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. In the future, we expect that emsReACT can also be used in actual EMS scenes. But further user studies are required to improve the performance of emsReACT where no instructors are present. In the future, the methods discussed in this research can be extended to address in-home emergency situations using existing systems such as Alexa, Google Home, etc. Our future goal includes using reinforcement learning instead of rule-based solutions for real-time assistance via safety-critical applications.

For the data mixing strategy and ablations in EMS-BERT, we do not have any ablation study at the moment to support the equal nature in core and subordinate corpora for augmenting the dataset. We have the data mixing research as a future goal for the project. Our future study will target finding what proportion of mixing both kinds of corpora yield best results, and whether there exist other approaches for data augmentation with similar or better results. Different methods for augmenting a dataset with amplified vocabulary exist in the literature, such as LSTM and transfer learning based approaches. As a future milestone of this research, we will investigate other data augmentation methods, run more comprehensive ablation studies for simultaneous pre-training, and compare their results with our current approach. As we are focused to develop a generic model to detect EMS concepts and understand their correlations, we concentrated on developing a custom BERT for EMS domain. Our future goal also includes using EMS-BERT for other down stream tasks such as negation detection, vitals validation, etc. from EMS corpora.

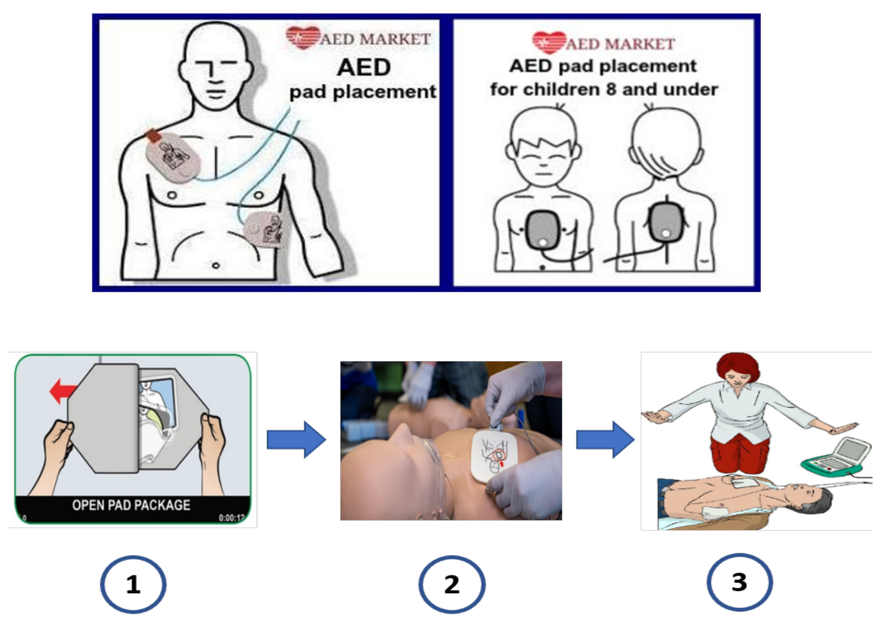


Figure 7.1: Basic types and gestures for CPR

The SenseEMS results currently highlight the basic gesture detection for CPR compressions only. In the future, we will combine the gestures for continuous activity detection from continuous sensor stream for CPR compressions, and other hand operated interventions. Besides the rate of CPR, we also aim to deduce depth of compressions in real-time. Currently, we are working on distinguishing the gestures for defibrillation (Defib) pad and bag-valve-mask (BVM) attachments, as shown in Figure 7.1.

Future work for EgoCap includes creating an EMS dataset using first-person captioning, and developing an enhanced transformer network to fuse the contextual knowledge which brings about state-of-the-art captioning on the EgoCap dataset. Currently, we are working on developing an enhanced transformer network that fuses the contextual knowledge using a stacked, multi-headed cross-attention layer alongside visual features.

Bibliography

Say, Bilge and Varol Akman (1996). "Current approaches to punctuation in computational linguistics". In: *Computers and the Humanities* 30.6, pp. 457–469.

Chapman, Wendy W et al. (2001). "Evaluation of negation phrases in narrative clinical reports." In: *Proceedings of the AMIA Symposium*. American Medical Informatics Association, p. 105.

McEntyre, Johanna and David Lipman (2001). "PubMed: bridging the information gap". In: *Cmaj* 164.9, pp. 1317–1319.

Papineni, Kishore et al. (2002). "BLEU: a method for automatic evaluation of machine translation". In: *Proceedings of the 40th annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, pp. 311–318.

Pollack, Martha E et al. (2002). "Pearl: A mobile robotic assistant for the elderly". In: *AAAI workshop on automation as eldercare*. Vol. 2002. AAAI, 2002, Edmonton, Alberta, Canada, pp. 85–91.

Pozner, Charles N et al. (2004). "International EMS systems: the United States: past, present, and future". In: *Resuscitation* 60.3, pp. 239–244.

Finkel, Jenny Rose, Trond Grenager, and Christopher Manning (2005). "Incorporating non-local information into information extraction systems by gibbs sampling". In: *Proceedings of the 43rd annual meeting on association for computational linguistics*. Association for Computational Linguistics, pp. 363–370.

LaFleur-Brooks, Myrna and Danielle S LaFleur (2005). *Exploring medical language: A student-directed approach*. Elsevier Mosby.

Müller, Dirk, Rahul Agrawal, and Hans-Richard Arntz (2006). “How sudden is sudden cardiac death?” In: *Circulation* 114.11, pp. 1146–1150.

Agarwal, Sumeet et al. (2007). “How much noise is too much: A study in automatic text classification”. In: *Seventh IEEE International Conference on Data Mining (ICDM 2007)*. IEEE, pp. 3–12.

Daily, Bonnie F, James W Bishop, Robert Steiner, et al. (2007). “The mediating role of EMS teamwork as it pertains to HR factors and perceived environmental performance”. In: *Journal of Applied Business Research (JABR)* 23.1.

Gamberini, Luciano et al. (2009). “Eldergames project: An innovative mixed reality table-top solution to preserve cognitive functions in elderly people”. In: *2009 2nd Conference on Human System Interactions*. IEEE, pp. 164–169.

Subramaniam, L Venkata et al. (2009). “A survey of types of text noise and techniques to handle noisy text”. In: *Proceedings of The Third Workshop on Analytics for Noisy Unstructured Text Data*. ACM, pp. 115–122.

Aronson, Alan R and François-Michel Lang (2010). “An overview of MetaMap: historical perspective and recent advances”. In: *Journal of the American Medical Informatics Association* 17.3, pp. 229–236.

Cer, Daniel M et al. (2010). “Parsing to Stanford Dependencies: Trade-offs between Speed and Accuracy.” In: *LREC*. Floriana, Malta.

Savova, Guergana K et al. (2010). “Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications”. In: *Journal of the American Medical Informatics Association* 17.5, pp. 507–513.

Burnett, Susan J et al. (2011). “Missing clinical information in NHS hospital outpatient clinics: prevalence, causes and effects on patient care”. In: *BMC health services research* 11.1, p. 114.

Chen, David L. and William B. Dolan (2011). "Collecting Highly Parallel Data for Paraphrase Evaluation". In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics)*, pp. 190–200.

Hu, John et al. (2011). "An advanced medical robotic system augmenting healthcare capabilities-robotic nursing assistant". In: *2011 IEEE international conference on robotics and automation*. IEEE, pp. 6264–6269.

Chang, Yao-Jen, Ya-Shu Kang, and Po-Chiao Huang (2013). "An augmented reality (AR)-based vocational task prompting system for people with cognitive impairments". In: *Research in developmental disabilities* 34.10, pp. 3049–3056.

DeVault, David et al. (2014). "SimSensei Kiosk: A virtual human interviewer for healthcare decision support". In: *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pp. 1061–1068.

Ha, Kiryong et al. (2014). "Towards wearable cognitive assistance". In: *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. ACM, pp. 68–81.

Hannun, Awni et al. (2014). "Deep speech: Scaling up end-to-end speech recognition". In: *arXiv preprint arXiv:1412.5567*.

Lin, Tsung-Yi et al. (2014). *Microsoft COCO: Common Objects in Context*. cite arxiv:1405.0312Comment: 1) updated annotation pipeline description and figures; 2) added new section describing datasets splits; 3) updated author list. URL: http://arxiv.org/abs/1405.0312.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning (2014). "Glove: Global vectors for word representation". In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.

Wu, Po-Liang et al. (2014). "A treatment validation protocol for cyber-physical-human medical systems". In: *2014 40th EUROMICRO Conference on Software Engineering and Advanced Applications*. IEEE, pp. 183–190.

Devarakonda, Murthy and Ching-Huei Tsou (2015). "Automated problem list generation from electronic medical records in IBM Watson". In: *Twenty-Seventh IAAI Conference*.

Graham, Robert, Margaret A McCoy, Andrea M Schultz, et al. (2015). "Understanding the public health burden of cardiac arrest: the need for national surveillance". In: *Strategies to Improve Cardiac Arrest Survival: A Time to Act*. National Academies Press (US).

Li, Ruijiao, Bowen Lu, and Klaus D McDonald-Maier (2015). "Cognitive assisted living ambient system: A survey". In: *Digital Communications and Networks* 1.4, pp. 229–252.

Lindes, Peter, Deryle W Lonsdale, and David W Embley (2015). "Ontology-based information extraction with a cognitive agent". In: *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Mehrabi, Saeed et al. (2015). "DEEPEN: A negation detection system for clinical text incorporating dependency relation into NegEx". In: *Journal of biomedical informatics* 54, pp. 213–219.

Rizzo, Albert et al. (2015). "Virtual reality standardized patients for clinical training". In: *The digital patient: Advancing healthcare, research, and education*, pp. 257–272.

Sennrich, Rico, Barry Haddow, and Alexandra Birch (2015). "Neural machine translation of rare words with subword units". In: *arXiv preprint arXiv:1508.07909*.

Sonntag, Daniel (2015). “Kognit: Intelligent cognitive enhancement technology by cognitive models and mixed reality for dementia patients”. In: *2015 AAAI Fall Symposium Series*.

Fan, Chenyou and David Crandall (2016). “DeepDiary: Automatically Captioning Lifelogging Image Streams”. In: *European Conference on Computer Vision (ECCV)*.

Guo, Chunhui, Shangping Ren, et al. (2016). “Transforming medical best practice guidelines to executable and verifiable statechart models”. In: *2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, pp. 1–10.

He, Yulan (2016). “Extracting topical phrases from clinical documents”. In: *Thirtieth AAAI Conference on Artificial Intelligence*.

Johnson, Alistair EW et al. (2016). “MIMIC-III, a freely accessible critical care database”. In: *Scientific data* 3.1, pp. 1–9.

Rajanna, Vijay et al. (2016). “KinoHaptics: An automated, wearable, Haptic assisted, physio-therapeutic system for post-surgery rehabilitation and self-care”. In: *Journal of medical systems* 40, pp. 1–12.

Tilk, Ottokar and Tanel Alumäe (2016). “Bidirectional Recurrent Neural Network with Attention Mechanism for Punctuation Restoration.” In: *Interspeech*, pp. 3047–3051.

Wen, Hongyi, Julian Ramos Rojas, and Anind K Dey (2016). “Serendipity: Finger gesture recognition using an off-the-shelf smartwatch”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 3847–3851.

Xu, Jun et al. (June 2016). “MSR-VTT: A Large Video Description Dataset for Bridging Video and Language”. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Bolanos, Marc et al. (Apr. 2017). "Egocentric Video Description based on Temporally-Linked Sequences". In: *Journal of Visual Communication and Image Representation* 50. DOI: 10.1016/j.jvcir.2017.11.022.

Bolaños, Marc, Mariella Dimiccoli, and Petia Radeva (2017). "Toward Storytelling From Visual Lifelogging: An Overview". In: *IEEE Transaction on Human-Machine Systems* 47.1, pp. 77–90. DOI: 10.1109/THMS.2016.2616296. URL: https://doi.org/10.1109/THMS.2016.2616296.

Guo, Chunhui, Zhicheng Fu, et al. (2017). "Pattern-based statechart modeling approach for medical best practice guidelines-a case study". In: *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, pp. 117–122.

Nicholson, Benjamin et al. (2017). "Emergency medical services (EMS) training in Kenya: Findings and recommendations from an educational assessment". In: *African Journal of Emergency Medicine* 7.4, pp. 157–159.

Shickel, Benjamin et al. (2017). "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis". In: *IEEE journal of biomedical and health informatics* 22.5, pp. 1589–1604.

Soysal, Ergin et al. (2017). "CLAMP–a toolkit for efficiently building customized clinical natural language processing pipelines". In: *Journal of the American Medical Informatics Association* 25.3, pp. 331–336.

Vaswani, Ashish et al. (2017). "Attention is all you need". In: *Advances in neural information processing systems* 30.

Anderson, P. et al. (2018). "Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering". In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Damen, Dima et al. (2018). “Scaling Egocentric Vision: The EPIC-KITCHENS Dataset”. In: *European Conference on Computer Vision (ECCV)*.

Devlin, Jacob et al. (2018). “Bert: Pre-training of deep bidirectional transformers for language understanding”. In: *arXiv preprint arXiv:1810.04805*.

Fan, Chenyou, Zehua Zhang, and David Crandall (Aug. 2018). “Deepdiary: Lifelogging image captioning and summarization”. In: *Journal of Visual Communication and Image Representation* 55, pp. 40–55.

He, Weipeng, Petr Motlicek, and Jean-Marc Odobez (2018). “Deep neural networks for multiple speaker detection and localization”. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 74–79.

Howard, Jeremy and Sebastian Ruder (2018). “Fine-tuned language models for text classification”. In: *arXiv preprint arXiv:1801.06146* 194.

Mondol, Md Abu Sayeed et al. (2018). “WaDa: An Android Smart Watch App for Sensor Data Collection”. In: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pp. 404–407.

Parish, David C, Hemant Goyal, and Francis C Dane (2018). “Mechanism of death: there’s more to it than sudden cardiac arrest”. In: *Journal of Thoracic Disease* 10.5, p. 3081.

Peters, Matthew E, Mark Neumann, Mohit Iyyer, et al. (2018). “Deep contextualized word representations”. In: *https://arxiv.org/abs/1802.05365v1*.

Poplin, Ryan et al. (2018). “Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning”. In: *Nature Biomedical Engineering* 2.3, pp. 158–164.

Preum, Sarah Masud, Sile Shu, Jonathan Ting, et al. (2018). “Towards a cognitive assistant system for emergency response”. In: *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems*.

Qian, Long, Anton Deguet, and Peter Kazanzides (2018). “ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery”. In: *Healthcare technology letters* 5.5, pp. 194–200.

Seki, Hiroshi et al. (2018). “A purely end-to-end system for multi-speaker speech recognition”. In: *arXiv preprint arXiv:1805.05826*.

Sharma, Piyush et al. (2018). “Conceptual Captions: A Cleaned, Hypernymed, Image Alt-text Dataset For Automatic Image Captioning”. In: *Proceedings of Association for Computational Linguistics*.

Sigurdsson, Gunnar A. et al. (June 2018). “Actor and Observer: Joint Modeling of First and Third-Person Videos”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, Utah, United States: IEEE, pp. 7396–7404. DOI: 10.1109/CVPR.2018.00772. URL: https://hal.inria.fr/hal-01755547.

Xiao, Cao, Edward Choi, and Jimeng Sun (2018). “Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review”. In: *Journal of the American Medical Informatics Association* 25.10, pp. 1419–1428.

Yu, Kun-Hsing, Andrew L Beam, and Isaac S Kohane (2018). “Artificial intelligence in healthcare”. In: *Nature biomedical engineering* 2.10, pp. 719–731.

Zhu, Peide et al. (2018). “Control with gestures: A hand gesture recognition system using off-the-shelf smartwatch”. In: *2018 4th International Conference on Big Data Computing and Communications (BIGCOM)*. IEEE, pp. 72–77.

Alsentzer, Emily et al. (2019). “Publicly available clinical BERT embeddings”. In: *arXiv preprint arXiv:1904.03323*.

Ardila, Diego et al. (2019). “End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography”. In: *Nature medicine* 25.6, pp. 954–961.

Chen, Zeya et al. (2019). “ARASID: Artificial Reverberation-Adjusted Indoor Speaker Identification Dealing with Variable Distances.” In: *EWSN*, pp. 154–165.

Huang, Kexin, Jaan Altosaar, and Rajesh Ranganath (2019). “Clinicalbert: Modeling clinical notes and predicting hospital readmission”. In: *arXiv preprint arXiv:1904.05342*.

Koutitas, George, Kenneth Scott Smith, et al. (2019). “A virtual and augmented reality platform for the training of first responders of the ambulance bus”. In: *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, pp. 299–302.

Montagna, Sara et al. (2019). “Real-time tracking and documentation in trauma management”. In: *Health informatics journal*, p. 1460458219825507.

Mujjiga, Srikanth et al. (2019). “Identifying Semantics in Clinical Reports Using Neural Machine Translation”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33, pp. 9552–9557.

Narayan, Sanjiv M, Paul J Wang, and James P Daubert (2019). “New concepts in sudden cardiac arrest to address an intractable epidemic: JACC state-of-the-art review”. In: *Journal of the American College of Cardiology* 73.1, pp. 70–88.

Pappagari, Raghavendra et al. (2019). “Hierarchical transformers for long document classification”. In: *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, pp. 838–844.

Peng, Yifan, Shankai Yan, and Zhiyong Lu (2019). “Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on ten benchmarking datasets”. In: *arXiv preprint arXiv:1906.05474*.

Peters, Matthew E, Mark Neumann, Robert L Logan IV, et al. (2019). “Knowledge enhanced contextual word representations”. In: *arXiv preprint arXiv:1909.04164*.

Preum, Sarah et al. (Aug. 2019a). “CognitiveEMS: A Cognitive Assistant System for Emergency Medical Services”. In: *SIGBED Rev.* 16.2, pp. 51–60. DOI: 10.1145/3357495.3357502. URL: https://doi.org/10.1145/3357495.3357502.

— (2019b). “CognitiveEMS: A cognitive assistant system for emergency medical services”. In: *ACM SIGBED Review* 16.2, pp. 51–60.

Rincon, Jaime A et al. (2019). “A new emotional robot assistant that facilitates human interaction and persuasion”. In: *Knowledge and Information Systems* 60, pp. 363–383.

Shu, Sile, Sarah Preum, Haydon M Pitchford, et al. (Aug. 2019). “A Behavior Tree Cognitive Assistant System for Emergency Medical Services”. In: *Intelligent Robots and Systems (IROS)*.

Shu, Sile, Sarah Preum, Haydon M Pitchford, et al. (2019). “A behavior tree cognitive assistant system for emergency medical services”. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 6188–6195.

You, Yang et al. (2019). “Large batch optimization for deep learning: Training bert in 76 minutes”. In: *arXiv preprint arXiv:1904.00962*.

Zhao, Sanqiang et al. (Jan. 2019). “Informative Image Captioning with External Sources of Information”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 6485–6494. DOI: 10.18653/v1/P19-1650.

Hobbs, Jonathan (2020). "It is time to rethink EMS". In: *The Veterinary Record* 187.9, pp. 363–364.

Koutitas, George, Scott Smith, and Grayson Lawrence (2020). "Performance evaluation of AR/VR training technologies for EMS first responders". In: *Virtual Reality*, pp. 1–12.

Lebanoff, Logan et al. (Nov. 2020). "Learning to Fuse Sentences with Transformers for Summarization". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, pp. 4136–4142. DOI: 10.18653/v1/2020.emnlp-main.338. URL: https://www.aclweb.org/anthology/2020.emnlp-main.338.

Lee, Jinhyuk et al. (2020). "BioBERT: a pre-trained biomedical language representation model for biomedical text mining". In: *Bioinformatics* 36.4, pp. 1234–1240.

Li, Yikuan et al. (2020). "BEHRT: transformer for electronic health records". In: *Scientific reports* 10.1, pp. 1–12.

Montagna, Sara et al. (2020). "Real-time tracking and documentation in trauma management". In: *Health informatics journal* 26.1, pp. 328–341.

Preum, Sarah Masud, Sile Shu, Homa Alemzadeh, et al. (2020). "Emscontext: EMS protocol-driven concept extraction for cognitive assistance in emergency response". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 08, pp. 13350–13355.

Rahman, M Arif, Sarah Preum, et al. (2020). "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.

Rahman, M Arif, Sarah M Preum, et al. (2020). "GRACE: generating summary reports automatically for cognitive assistance in emergency response". In: *Proceed-*

ings of the AAAI Conference on Artificial Intelligence. Vol. 34. 08, pp. 13356–13362.

Wada, Shoya et al. (2020). “Pre-training technique to localize medical bert and enhance biomedical bert”. In: *arXiv preprint arXiv:2005.07202*.

Wang, Yanshan et al. (2020). “MedSTS: a resource for clinical semantic textual similarity”. In: *Language Resources and Evaluation* 54.1, pp. 57–72.

Yao, Shuochao et al. (2020). “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)*. IEEE, pp. 1–10.

Al Amiry, Alaa and Brian J Maguire (2021). “Emergency medical services (Ems) calls during covid-19: Early lessons learned for systems planning (a narrative review)”. In: *Open Access Emergency Medicine: OAEM* 13, p. 407.

al., Kristen Grauman et. (2021). *Ego4D: Around the World in 3,000 Hours of Egocentric Video*. arXiv: 2110.07058 [cs.CV].

Bi, Hongliang et al. (2021). “Magic-hand: Turn a smartwatch into a mouse”. In: *Pervasive and Mobile Computing* 78, p. 101487.

Kim, Sion et al. (2021). “Information Extraction from Patient Care Reports for Intelligent Emergency Medical Services”. In: *2021 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, pp. 58–69.

Kolenik, Tine and Matjaž Gams (2021). “Intelligent cognitive assistants for attitude and behavior change support in mental health: state-of-the-art technical review”. In: *Electronics* 10.11, p. 1250.

Liu, W. et al. (2021). “CPTR: Full Transformer Network for Image Captioning”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Nguyen-Duc, Thanh et al. (2021). "Deep EHR Spotlight: a Framework and Mechanism to Highlight Events in Electronic Health Records for Explainable Predictions". In: *AMIA Summits on Translational Science Proceedings* 2021, p. 475.

Preum, Sarah Masud, Sirajum Munir, et al. (2021). "A review of cognitive assistants for healthcare: Trends, prospects, and future directions". In: *ACM Computing Surveys (CSUR)* 53.6, pp. 1–37.

Samyoun, Sirat et al. (2021). "iWash: A smartwatch handwashing quality assessment and reminder system with real-time feedback in the context of infectious disease". In: *Smart Health* 19, p. 100171.

Wang, Han et al. (2021). "A Weakly-Supervised Named Entity Recognition Machine Learning Approach for Emergency Medical Services Clinical Audit". In: *International journal of environmental research and public health* 18.15, p. 7776.

Kunwar, Utkarsh et al. (2022). "Robust and Deployable Gesture Recognition for Smartwatches". In: *27th International Conference on Intelligent User Interfaces*, pp. 277–291.

Appendices

Appendix A

Appendix - GRACE: Generating Summary Reports Automatically for Cognitive Assistance in Emergency Response

A.1 Detailed Results for All Fields of The Form

Table A.1: Detailed accuracy of miCaRe for all fields of our model summary report form for different data types

Field in the Form	Accuracy Measure	Noise-free audio	Noisy audio	Noise-free narratives	Noisy narratives
Demographic Information	Precision	0.97	0.97	1.00	0.94
	Recall	0.85	0.82	1.00	0.93
	F1 Score	0.91	0.89	1.00	0.93
Chief Complaint (CC)	Precision	0.88	0.77	0.95	0.83
	Recall	0.81	0.67	0.91	0.79
	F1 Score	0.85	0.72	0.93	0.81
History of Present Illness (HPI)	Precision	0.81	0.75	0.85	0.81
	Recall	0.57	0.51	0.77	0.64
	F1 Score	0.67	0.61	0.81	0.72
Past Medical History (PMH)	Precision	0.61	0.42	0.84	0.72
	Recall	0.89	0.70	0.78	0.71
	F1 Score	0.73	0.53	0.81	0.72
Allergies	Precision	0.75	0.69	0.81	0.69
	Recall	0.61	0.61	0.70	0.61
	F1 Score	0.68	0.65	0.76	0.65
Medication (Meds)	Precision	0.87	0.81	0.92	0.78
	Recall	0.91	0.76	0.81	0.71
	F1 score	0.89	0.79	0.87	0.75
Procedure/Treatment/Transport (PE/RX/TX)	Precision	0.87	0.72	0.84	0.65
	Recall	0.81	0.70	0.78	0.72
	F1 Score	0.84	0.71	0.81	0.69
Vital Signs	Precision	0.98	0.88	0.97	0.84
	Recall	0.95	0.83	0.96	0.81
	F1 Score	0.967	0.86	0.97	0.83
Procedure Details	Precision	0.83	0.72	0.89	0.70
	Recall	0.81	0.62	0.77	0.53
	F1 Score	0.82	0.67	0.83	0.61
Medication Administrated	Precision	0.87	0.71	0.94	0.84
	Recall	0.98	0.90	0.87	0.79
	F1 Score	0.93	0.80	0.91	0.82

Appendix B

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

B.1 Survey Details

31 responders took this survey for evaluating miCaRe, none of these responders were involved in the development phase. Due to the ongoing pandemic, we could not meet the responders, so they did not test miCaRe physically. However we shared a detailed video demonstrating the use of miCaRe during simulated training scenes. Following Likert scale based rating questionnaire was used to understand what the responders thought about miCaRe, the responders provided additional background information such as their experience level, professional span, detailed list of specialized interventions, etc. We also collected open ended answers to specific questions, we include some of those questions here.

(i) Technology can provide crucial assistance when combined with human efforts during EMS scenes.

- Strongly Disagree • Disagree • Agree • Strongly Agree • Neither Agree or Disagree

(ii) Using electronic devices and gadgets such as microphone and smartwatch during EMS scene does not hinder the process of providing care.

- Strongly Disagree • Disagree • Agree • Strongly Agree • Neither Agree or Disagree

(iii) If you answered Agree or Strongly Agree in the question above, what other devices you feel might be used during EMS scene without adding any burden on the responder? For example, smart glasses, drones, robots, etc.

- Your Answer:

(iv) The responders will not require too much adaptation in their course of action if miCaRe is used during real EMS scenes.

- Strongly Disagree • Disagree • Agree • Strongly Agree • Neither Agree or Disagree

(v) If you answered Disagree or Strongly Disagree in the question above, why do you think the adaptation will be difficult?

- Your Answer:

(vi) Even though EMS scenes are dynamic and fast evolving, miCaRe assistant can keep up with the responders and will not slow down the care providing process.

- Strongly Disagree • Disagree • Agree • Strongly Agree • Neither Agree or Disagree

(vii) miCaRe will help standardize EMS protocols over broader geographic territory and provide a common platform for training.

- Strongly Disagree • Disagree • Agree • Strongly Agree • Neither Agree or Disagree

(viii) Using conversational audio data for miCaRe is effective for EMS training and real EMS scenes.

- Strongly Disagree • Disagree • Agree • Strongly Agree • Neither Agree or Disagree

(ix) Using sensor data from electronic devices for miCaRe is effective for EMS training and real EMS scenes.

- Strongly Disagree • Disagree • Agree • Strongly Agree • Neither Agree or Disagree

(x) Interactive feedback feature of miCaRe is effective for EMS training and real EMS scenes.

- Strongly Disagree • Disagree • Agree • Strongly Agree • Neither Agree or Disagree

(xi) Post scene quality assurance of CPR of miCaRe is effective for EMS training and real EMS scenes.

- Strongly Disagree • Disagree • Agree • Strongly Agree • Neither Agree or Disagree

(xii) Using miCaRe during EMS training and real EMS scenes will be helpful for the responders.

- Strongly Disagree • Disagree • Agree • Strongly Agree • Neither Agree or Disagree

(xiii) Overall, the idea and performance of miCaRe in EMS training scenes and potential real scenes in future will be-

- Problematic • Below Standard • Standard • Useful • Very Useful

B.2 Example of a training scene conversation among responders

Here, provider 1 is registered as a paramedic, provider 2 is an EMT-basic, provider 3 is a firefighter. The less the expertise level, the more detailed the feedback from miCaRe.

Provider 1: Unit 94 on scene.

Provider 3: Mr. John Doe is in cardiac arrest. Starting compressions now.

emsReACT Log Display: Patient Name= John Doe. Time of compression initiation 12:02:00. Timer set for CPR provider switch every 5 minutes.

Provider 2: I'm going to start BVM with 6 Litres of Oxygen.

emsReACT Log Display: Time of BVM initiation = 12:02:10. Parameters 6 Litres of Oxygen.

emsReACT Emergency Prompt: **Please check for chest rise.**

Provider 3: There's good chest rise.

emsReACT Log Display: Good chest rise at 12:02:15.

Provider 1: Defibrillation pads attached. The patient is in ventricular fibrillation.

emsReACT Log Display: Time of attachment of defibrillation pads = 12:02:30.

emsReACT Emergency Prompt: **Start defibrillation at 120 Joules. Please check for EKG waveform.**

Provider 1: Alright everyone, we're defibrillating at 120 Joules! Hands off.

emsReACT Log Display: Time of first defibrillation = 12:03:00. Energy level = 120 Joules. Timer started for defibrillation protocol for every 2 min.

Provider 1: No ROSC, I'm starting CPR again.

Provider 1: Okay, I'm going to try to start an IV on the left arm.

emsReACT Log Display: Started timer for vascular access protocol.

Provider 3: I'm going to try for intubation.

emsReACT Log Display:Please check if you are allowed to do this intervention! Time of first intubation attempt 12:03:30.

Provider 1: Okay I will try! Shoot, I can't get the IV.

Provider 1: Alright, airway is secured with 8.0 ET tube.

emsReACT Log Display: Time airway was achieved 12:04:00. Type and size of airway 8.0 ET.

emsReACT Emergency Prompt: **Please auscultate lungs, check chest rise, check vitals ETCO2.**

Provider 3: I'm seeing good chest rise and I'm hearing good lung sounds bilaterally.

emsReACT Log Display: Good chest rise, good lung sounds bilaterally at 12:04:20.

emsReACT Emergency Prompt: **Please second defibrillation at 150 Joules in 10 seconds at 12:05:00.**

Provider 2: Time to defibrillate at 150 Joules. Stop CPR!

emsReACT Log Display: Second defibrillation time = 12:05:00 Energy level = 150 Joules.

Provider 1: No ROSC again!

emsReACT Emergency Prompt: **Please restart CPR.**

emsReACT Emergency Prompt: **In 10 seconds start intraosseous access.**

Provider 1: CPR started. Okay, I'm going to go for an IO in the right humerus. IO is in place in the right humerus, starting normal saline, wide open. *emsReACT Log Display*: Time CPR: 12:07:30, access achieved = 12:07:30. Type of fluid started = normal saline. Location = right humerus.

emsReACT Emergency Prompt: **Please start epinephrine sequence protocol. Administer epinephrine 1:10,000 1mg IVP**

Provider 1: Okay, I'm giving Mr. Doe epinephrine 1:10,000 1mg IV push.

emsReACT Log Display: Medication name = epinephrine, dosage = 1:10,00 1mg, route of administration = IVP, time of administration = 12:08:00. Timer for epinephrine sequence every two minutes started.

emsReACT Emergency Prompt: **Please consider post-medication flush and Amiodarone 300 mg IVP.**

Provider 1: Giving the post medication flush and an IV push of Amiodarone 300 mg now.

emsReACT Log Display: Medication name = Amiodarone, dosage = 300 mg, route of administration = IVP, time of administration = 12:08:30.

emsReACT Emergency Prompt: **Please administer second epinephrine dosage in ten seconds at 12:10:00.**

Provider 1: Okay I'm giving epinephrine again now.

Provider 1: This is epinephrine 1:10,000 1mg given by IV push and we're giving it to Mr. Doe.

emsReACT Log Display: Medication name = epinephrine, dosage = 1:10,000 1mg, route of administration = IVP, time of administration = 12:10:00.

emsReACT Emergency Prompt: **Please switch CPR provider, 5 minutes up!**

Provider 3: Okay I'll take over CPR.

emsReACT Log Display: Full adult, follow CPR protocol one please. Time of provider change 12:12:30. Restarted CPR protocol timer.

Provider 2: Let's do a medication cross check. Show all medications applied.

emsReACT Log Display: 1. IO access achieved = 12:07:30. Type of fluid started = normal saline. Location = right humerus. Medication name = epinephrine, dosage = 1:10,00 1mg, route of administration = IVP, time of administration = 12:08:00.

2. Medication name = epinephrine, dosage = 1:10,00 1mg, route of administration = IVP, time of administration = 12:08:00.

3.Medication name = Amiodarone, dosage = 300 mg, route of administration = IVP, time of administration = 12:08:30. Medication name = epinephrine, dosage = 1:10,000 1mg, route of administration = IVP, time of administration = 12:10:20.

Provider 1: We have ROSC.

emsReACT Log Display: Time of ROSC found 12:17:00.

emsReACT Emergency Prompt: **Please administer amiodarone drip 150 mg over 10 mins by IV pump.**

Provider 1: How do you dilute the amiodarone again?

emsReACT Emergency Prompt: **Dilute 150mg of amiodarone in 100mL D5W to yield 1.5mg/mL.**

Provider 1: I'm administering amiodarone drip 150 mg over 10 minutes via IV pump.

emsReACT Log Display: Medication name = amiodarone, dosage = 150mg, route of administration = IV drip over 10 mins, time of administration = 12:18:00.

emsReACT Emergency Prompt: **Please consider post medication flush.**

Provider 1: Giving the post medication flush and an IV push of Amiodarone 300 mg now.

emsReACT Log Display: Medication name = Amiodarone, dosage = 300 mg, route of administration = IVP, time of administration = 12:18:30.

Provider 1: Okay we are done now, patient is responding let's transfer him to hospital. Generate log reports now. emsReACT Emergency Prompt: **Patient summary report generated at 12:20:00. Please check the activity log document to verify. Thanks, goodbye!**

Appendix C

Appendix - EMS-BERT: A Pre-Trained Language Representation Model for the Emergency Medical Services (EMS) Domain

C.1 A sample document from the EMS corpus

(De-identified) "*Dispatched for a sick person, that was changed to abdominal pain en route. Arrived to find the patient sitting outside of his house on porch steps leaning over holding his abdomen. Abdominal Pain found. The patient stated that he was discharged yesterday from local hospital and was diagnosed with pancreatitis. His pain onset was yesterday after dilaudid wore off, prescribed percocets did no effect with pain management per patient. Pain described as sharp pain, 10 out of 10, no radiation. He stated that his pain was in his entire abdomen, but worse in lower quadrants. The patient has had a heart attack, 2 stents put in. He stated that he has history of bradycardia. The patient was also given medication for nausea, but it did not work and* *he is still nauseous. PMH, medications, and medication allergies as noted. Alert and oriented person, place, event, and time are aware. Airway is patent, unobstructed, and self maintained. Breathing has equal bilateral chest rise and fall, clear lung sounds. Circulation has pms x 4, no bleeding noted, good turgor, no diaphoresis, skin warm pink and dry. The patient denies chest pain, shortness of breath, vomiting, back pain, syncope, lightheadedness, or blurry vision. Vitals obtained and documented. 12 lead: Sinus Bradycardia. 20 G IV in right hand connected to a 10gtts/ml drop set connected to a bag of 2000ml NS. 250ML Bolus of NS administered. Transported the patient to local hospital per patient's request. Patient condition unchanged. Report and care given to RN in room. Unable to obtain the patient's insurance due to his high level of pain.*"