Journal of Innovations in Computer Science and Engineering
Shahid Beheshti Unversity

June 2025, Volume 3, Issue 1

Persian Intelligent Assistant in Healthcare Domain

Sarina Chitsaz✉, **Code Orcide:** **0009-0008-8237-3658**
Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran, s.chitsaz@mail.sbu.ac.ir
Mehrnoush Shamsfard, **Code Orcide:** **0000-0002-7027-7529**
Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran, m-shams@sbu.ac.ir

Abstract—Nowadays, advances in technology and medical science have led to significant changes in the field of healthcare. Consequently, an effort has been taken to develop an intelligent health assistant in the Persian language, focusing on the emergency department. To achieve this goal, a labeled dataset was prepared. Subsequently, an intelligent assistant architecture was developed, utilizing slot filling and speech act classification for natural language understanding. A dialogue manager was designed to address negation in patient statements, resulting in the classification of triage patients. Evaluation revealed that the assistant's performance matched that of emergency staff in 83% of cases.

Keywords— Intelligent Assistant, Natural language understanding, Speech act classification, Slot filling

Introduction

Considering the importance and status of the health sector, the rapid changes in today's society and the increasing needs and expectations regarding health services, people want to receive higher quality of services. Additionally, taking into account the vital role of factors such as accuracy, speed and reliability in healthcare and the vulnerability of the health system to human errors, utilizing artificial intelligence can bring comfort to patients as well as promoting health indicators in society. It can be acknowledged that nowadays using intelligent systems in health sector has become a necessity. Employing assistants that imitate human interactions can significantly transform this field and greatly impact communication with patients and their companions. Intelligent healthcare helps in maximizing the potential of existing resources which helps in remote monitoring of patients and reducing the costs of treatment. It also allows specialist doctors to provide their services without any geographical barriers. Due to the increasing number of visiting patients in medical centers, the use of intelligent assistants leads to saving both time and cost. As a result, healthcare workers are free to do more complex tasks. Owing to the critical importance of time and accuracy for medical staff in saving patients' lives, especially in the emergency department, the use of these assistants will reduce response times for patients in critical situations. Additionally, it will increase accuracy and focus on high-risk patients, prioritizing their care.

On the other hand, the use of the above technology has decreased unnecessary visits to health care centers. Thus, performing daily repetitive activities including providing inpatient and outpatient services has become faster and more precise.

Triage is a prioritization system and a decision-making process to manage the admission of patients and provide services to them in the emergency department of the hospital.

The purpose of triage in the emergency department is to answer the question: "At this moment, what is the priority of care of this particular patient among all patients attending the emergency department?" [1]. One of the methods of triaging patients is the "Emergency Severity Index (ESI.V) [2]," which is considered and used by the National Hospital Emergency Triage Committee in Iran as the most appropriate triage algorithm. The triage system is a five-level, easy-to-use tool that categorizes emergency department patients by simultaneously examining the severity of the condition and its required intervention [3]. In the first phase, the triage nurse estimates the severity of the disease based on scientific medical principles, which are categorized into 5 levels as follows:

- Level 1: It includes patients who often do not have vital symptoms and have life-threatening conditions.
- Level 2: Level 2 patients include patients who need to be visited by an emergency medicine specialist within ten minutes.
- Level 3: Most of the patients referred to the emergency department of the hospital are at this level, and they are visited and examined by a specialist emergency medicine physician in less than an hour.
- Level 4: This level includes patients who do not have any life-threatening symptoms and are visited by a specialist doctor in less than two hours, depending on the level of crowding and the number of patients in the department.
- Level 5: This level includes patients who have come for different requests such as simple visits. The duration of providing services to these patients will be less than four hours.

If the severity of the disease is not high (level 1 and 2 of triage), the triage nurse determines the triage level (level three, four and five) by estimating the number of required facilities.

Although working in medical domain has its merits, there are some challenges related to the features of this area. Characteristics of medical texts include [4]: telegraphic style (incomplete sentences), short texts (abbreviations, special phrases and characters), and negation (negation words or phrases)

Submit Date: 2025-09-28
Accept Date: 2025-10-05
✉ Corresponding author

In light of the of the benefits and challenges mentioned earlier, and the research that have been done so far, a Persian intelligent assistant in the field of healthcare focused on the emergency has not been designed. This is why the present research can bring several benefits to this area.
The contributions of this paper are summarized as follows:

- To the best of current knowledge, this Persian healthcare intelligent assistant in the emergency department represents the first dialogue system in this field.
- An annotated dataset, for training healthcare intelligent assistant models is generated.
- A framework is introduced for communicating with patients and extracting maximum information from their statements in order to respond appropriately.
- The model's performance in predicting triage levels of patients is evaluated.

The structure of the next sections of the article is as follows: In the second section, a review of related research is presented. In the third section, the general structure of the proposed approach is described, and in the fourth section, the obtained results are analyzed. Finally, in the fifth section, a summary of the research is provided.

Related Work

In general, the architecture of intelligent assistants contains 4 parts [5]: natural language preprocessing, natural language understanding (NLU), dialogue manager, and natural language generation.
After receiving the user's input and performing the necessary pre-processing, the dialogue system checks the user's input [6]. A frame structure can be used to understand the user's request that consists of two tasks: intent detection and slot filling.
In recent work, Shi et al. [7] address medical slot filling as a multi-label classification problem using a label-embedding attention model. This model identifies which words are highly relevant to the slot-filling task and selects keywords accordingly. After that, they use a text classification encoder as the query encoder.
BERT [8] has also been used for slot filling and intent detection. One of these studies was conducted by Chen et al. [9], where BERT was employed to perform both tasks jointly. This model was implemented on a dataset of size 4478 with 120 slots and 21 intents and on another dataset of size 13084 with 72 slots and 7 intents.
Inou et al. [10] presented an intelligent triage system called Healthbot for emergency rooms. This system interacts with patients and categorizes them based on priority level and medical specialty. The system utilizes the Google Dialogflow conversation interface along with sensors, voice recognition and a physical robot-like platform. The classification of medical specialties in Healthbot is divided into four groups: general medicine, surgery, orthopedics, and cardiology. The accuracy of the specialty classification in this bot is reported to be 82.2%.
In recent research, the emergence of ChatGPT has enabled its use as an intelligent health assistant. In a study involving 56 cases, Gebrael et al. [11] used ChatGPT to provide emergency care for patients with metastatic prostate cancer. They employed ChatGPT to assess the need for hospitalization or discharge and acknowledged its high sensitivity in determining patient admissions.

Approach

The proposed approach involves two sections: dataset generation and model architecture development.

Dataset Generation

At the first step, a dataset comprising exact formal and informal dialogues was collected from the emergency department triage in a hospital. Second, the process of de-identification was conducted. This process involves cleaning and removing identifying information to ensure the privacy of individuals. Finally, the dataset was labeled for speech act classification and slot filling. In total, the dataset contains 5,231 utterances.
Speech Acts Dataset: The speech acts of emergency department visitors include categories of information expressed by speakers based on the topic of conversation. There are eight categories in this dataset, named:

- Patient Signs
- Patient Information, such as age
- Patient History, including medical history, family history, etc.
- Patient Requests, such as serum injection or suturing
- Patient Questions
- Patient Confirmation
- Patient Negation
- Others: utterances that do not fall into any of the above categories

Each utterance in this section may consist of more than one speech act.
Slots Dataset: The labels in this part define the type of semantic information conveyed by each word. The slot filling task is considered as a sequence labeling task, where a chain of text maps to a chain of I, O, and B tags. In this method, each token is classified into one of three tags, representing the token's status in the sequence in an organized structure:

- B (Beginning): Represents the start of a new slot in the structure. A token that begins a new entity is labeled as "B."
- I (Inside): Indicates that a token is part of a specific structural slot, continuing from a preceding "B" tag.
- O (Outside): Denotes tokens that do not belong to any structural slots and are outside of all defined entities.

Once the sequence labeler has tagged a user utterance, a filler string for each slot in the tags can be extracted. The system can then decide how to complete the task for the user, ensuring precise dataset labeling. Owing to this, labeling was carried out in consultation with several experts in the field of healthcare. In addition to the "O" tag, 61 tags were defined within the dataset.
An example of Slot Labels and Speech Act Labels is shown in Fig.1.

Speech Act Labels	Slot Labels			
	دارد	بینی	آبریزش	پسرم
Patient Signs Patient Information	O	I_SYMPTOM	B_SYMPTOM	B_GENDER

Speech Act Labels	Slot Labels					
	My	son	has	a	runny	nose
Patient Signs Patient Information	O	B_GENDER	O	O	B_SYMPTOM	I_SYMPTOM

Fig. 1. Slot Labels and Speech Act Labels

Proposed Model Architecture

At the start of the intelligent assistant's work, after receiving the patient’s utterance, the input is sent to the natural language understanding module to comprehend what has been expressed. The NLU module is designed to enable computers to interpret human language in a way that is both meaningful and contextually relevant, converting human language into a machine-understandable structure. In this module, preprocessing is performed on the input, and the existing speech acts, along with the values of predetermined slots, are identified.

Speech act classification was treated as a text classification task. For this purpose, a BERT [8] pre-trained language model, one of the state-of-the-art methods in natural language understanding, was used. BERT is a transformer-based model initially trained on massive datasets such as Wikipedia and BooksCorpus. It also, was employed for the slot-filling task. Regarding the dataset's language is Persian, the tasks were fine-tuned on ParsBERT [12], a BERT-based model specifically designed for processing Persian text. For the NLU module, a joint framework for speech act classification and slot filling was proposed.

After understanding the utterance of the patient it is time for Dialogue manager. This module has a crucial role in controlling and management of the dialogue flow between a patient and the assistant. Dialogue manager consists of the Dialogue State Tracker (DST) and the Policy Manager. DST maintains the current status of a conversation while Policy Manager has the responsibility of managing and coordinating various policies in the intelligent assistant. It determines the system’s reactions and responses to the patient in specific situations. The defined policies in this module are a combination of rule-based and machine learning techniques.

When the speech acts and slot fillers are detected, the DST is updated and the new information is added to the previous data. Then the policies of the triage level 1 and level 2 are analyzed. The policies of the intelligent assistant are designed based on the protocols of emergency triage in Iran where uses ESI algorithm [3]. In every turn of the dialogue, if the patient’s condition is identified as level 1 or 2, immediate action is required. Otherwise, the prioritized questions for getting other information and symptoms of the patient will be asked by the assistant. The above procedure will repeat until all the required information is obtained.

Consequently, if the gathered information is sufficient, the triage level 3, 4 and five policies will check and the triage level of the patient will be detected. In the following, essential actions and guidance will be presented according to the patient's condition. Additionally, in this step, a patient's medical history has been achieved on the account of the dialogue state tracker and the saved slots that hold information, signs, and the patient's history.

It is worth mentioning that one of the main challenges in the interaction between the patient and the intelligent assistant is the existence of negation in utterances.

Negation Detection Approach: During conversations, a patient can give positive or negative answers to the questions of the assistant. These replies can be divided into “Patient Confirmation” and “Patient Negation” by speech act classification which either of them have their own suitable policy.

Negation changes the entire meaning of an utterance, and its existence makes it one of the important challenges in the healthcare field. In this research, the designed approach for the Negation Detection task has been developed by combining deep learning and rule-based techniques based on the Persian language.

As Rahimi and Shamsfard stated in their article [13], due to differences between the structural patterns of negation in the Persian language and the English language, the available methods must be modified and properly adapted. One example of these differences is given below:

- When using negative adverbs like “never (هرگز)” in the sentence “I have never smoked,” the English pattern uses a negative adverb with a positive verb. However, in the Persian pattern, as in the sentence “هرگز سیگار نکشیده‌ام,” a negative verb is paired with a negative adverb.

Regarding these differences, the proposed approach uses a negation tag for negative verbs. This label is accessible in the Slots Dataset. In this way, negation can be detected simultaneously along with other slots by the ParsBERT transformer model.

On the other hand, there is an exception. In the Slots dataset, sentences like “I have no balance (تعادل ندارم)” are classified under the “Patient Signs” tag instead of “Patient Negation.” As a result, for detecting negation in the text, a two-stage mechanism is introduced. In this mechanism, if the output of the NLU module includes both the negation slot label and the speech act “Patient Negation” at the same time, negation is recognized in the text.

After detecting negation with the help of ParsBERT, the next step is Negation Scope Resolution [14], which is the process of finding the negated part of a sentence. Inspired by the rule-based methods of NegEx [15] and ConText [16], and considering the Persian language, the previously extracted slot before the negative sign is affected and becomes negated. At this point in the intelligent assistant, the negation sign, along with its corresponding slot, is removed from the DST to only consider the patient's existing condition.

Experiments

The assistant's performance has been evaluated in the classification of triage patients.
The proposed model was implemented in Python programming language and the Google Colab environment, utilizing a Tesla T4 GPU with 16 GB of memory. The proposed NLU model was fine-tuned on ParsBert using the introduced dataset. After conducting several tests, hyperparameters such as a batch size of 32, a learning rate of 2e-5, and a maximum length of 64 were used, in addition to Adam's optimizer [17] and early stopping. Considering the fact that the assistant operates in the health-care domain, the 5-layer cross-validation method has been used for more reliable results.
Dataset: According to the nature of this assistant, to evaluate the correctness of its diagnosis and answers, the test dataset was provided. This dataset was developed under the supervision of three emergency medicine specialists, including the following features:

- It consists of 132 case scenarios.
- 14 cases did not meet the input criteria.
- 96 cases were low-risk patients (level 3, 4 and 5 ESI triage)

Evaluation Metrics: For evaluating the performance of the intelligent assistant two metrics (Accuracy and F1-Score) have been used.

Accuracy: In the context of a triage system like ESI, accuracy measures how well the system classifies patients into their relevant severity levels.

F1-Score: F1-Score is based on true positive, true negative, false positive and false negative values. Each of these is defined in the triage system as follows:

- True positive: The system correctly identifies the level of patient triage as high severity and it is also high in reality.
- True negative: The system correctly identifies the level of patient triage at low severity and it is also low in reality.
- False Positive: The system incorrectly identifies a patient's triage level as high severity while it is not high in reality (Over-Triage).
- False negative: The system incorrectly identifies a patient's triage level as low severity while it is high in reality (Under-Triage).

Results: In this section, the values of Accuracy and F1-Score were calculated as 83% and 89.32%, respectively. The confusion matrix of the intelligent healthcare assistant's predictions and the reference's triage levels is shown in Table I.

TABLE I. CONFUSION MATRIX OF DIFFERENT TRIAGE LEVELS (ROWS: ACTUAL LEVELS, COLUMNS: PREDICTED LEVELS)

Triage Levels	**1**	**2**	**3**	**4**	**5**	**Total**
1	7	1	0	0	0	8
2	1	10	3	0	0	14
3	0	2	48	5	1	56
4	0	0	2	19	3	24
5	0	0	0	2	14	16
Total	8	13	53	26	18	118

In the classification of ESI triage, determining the exact level of the patient is vital. As shown in the confusion matrix in Table I, the intelligent assistant correctly classified the majority of cases. The results are discussed below:

- Out of 8 patient cases at level 1, the intelligent assistant correctly identified the level in 7 cases, and only in 1 case, the triage level was incorrectly declared as level 2.
- Out of 14 patient cases at level 2, the intelligent assistant correctly identified the level in 10 cases and declared 1 case as level 1, and 3 cases as triage level 3.
- Out of 56 patient cases at level 3, the intelligent assistant correctly identified the level in 48 cases and declared 2 cases as level 2, 5 cases as level 4, and 1 case as triage level 5.
- Out of 24 patient cases at level 4, the intelligent assistant correctly identified the level in 19 cases and declared 2 cases as level 3, and 3 cases as triage level 5.
- Out of 16 patient cases at level 5, the intelligent assistant correctly identified the level in 14 cases, and only in 2 cases, the triage level was incorrectly declared as level 4.

According to the aforementioned results, it is observed that mistakes occurred relatively more in the triage levels with greater categorical overlap.
Moreover, as shown in the statistics in Table I, the assistant's best performance was in determining Level 3, indicating that patients requiring multiple resources are accurately categorized, preventing unnecessary escalation to higher levels. In second place, Level 1 performed well, revealing that the model works effectively for patients who have life-threatening conditions and

need immediate interventions. After Level 1, in third place, the assistant was effective in Level 5 diagnosis, correctly identifying the majority of non-urgent patients, helping optimize resources.

In addition, it should be noted that the performance results of this assistant in determining the triage levels of patients demonstrate high diagnostic accuracy, and in 83% of cases, it was consistent with the performance of emergency staff.

Conclusion and Future Work

In this research, an intelligent assistant for the healthcare domain was developed. Its purpose was to assist the emergency department in prioritizing between high-risk and low-risk patients based on the ESI algorithm. The process commenced with data preparation through field research to collect data from hospitals. Subsequently, a joint model was used to fine-tune ParsBERT to understand patients' utterances with the help of a negation detection mechanism. Patients were then prioritized according to the extracted information and the defined policies related to the emergency triage algorithm. In the final stage, the performance of the assistant was evaluated using a dataset prepared under the supervision of experts in the relevant field. The results showed that, in the majority of cases (83%), the assistant correctly identified the patients' triage levels.

In future work, the dataset will be expanded to enhance the robustness and generalizability of the outcomes. Additionally, more evaluations will be conducted. One of them will be ablation studies of the negation module in order to investigate its contributions within the assistant. The other will be a comprehensive comparison (specifically in the context of few-shot learning capabilities) between the healthcare assistant and large language models (LLMs) to identify strengths, weaknesses, and areas of improvement in the reliability and performance of the healthcare intelligent assistant.

References

[1] R. Sánchez-Salmerón et al., "Machine learning methods applied to triage in emergency services: A systematic review," Int. Emerg. Nurs., vol. 60, p. 101109, Jan. 2022, doi: 10.1016/j.ienj.2021.101109.

[2] L.-H. Yao, K.-C. Leung, C.-L. Tsai, C.-H. Huang, and L.-C. Fu, "A Novel Deep Learning–Based System for Triage in the Emergency Department Using Electronic Medical Records: Retrospective Cohort Study," J. Med. Internet Res., vol. 23, no. 12, p. e27008, Dec. 2021, doi: 10.2196/27008.

[3] N. Shafaf and H. Malek, "Applications of Machine Learning Approaches in Emergency Medicine; a Review Article," Arch. Acad. Emerg. Med., vol. 7, no. 1, Art. no. 1, Jul. 2019, doi: 10.22037/aaem.v7i1.410.

[4] H. Lu, L. Ehwerhemuepha, and C. Rakovski, "A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance," BMC Med. Res. Methodol., vol. 22, no. 1, p. 181, Jul. 2022, doi: 10.1186/s12874-022-01665-y.

[5] L. Qin, T. Xie, W. Che, and T. Liu, "A Survey on Spoken Language Understanding: Recent Advances and New Frontiers," in Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Montreal, Canada: International Joint Conferences on Artificial Intelligence Organization, Aug. 2021, pp. 4577–4584. doi: 10.24963/ijcai.2021/622.

[6] H. Weld, X. Huang, S. Long, J. Poon, and S. C. Han, "A Survey of Joint Intent Detection and Slot Filling Models in Natural Language Understanding," ACM Comput. Surv., vol. 55, no. 8, p. 156:1-156:38, Dec. 2022, doi: 10.1145/3547138.

[7] X. Shi, H. Hu, W. Che, Z. Sun, T. Liu, and J. Huang, "Understanding Medical Conversations with Scattered Keyword Attention and Weak Supervision from Responses," Proc. AAAI Conf. Artif. Intell., vol. 34, no. 05, Art. no. 05, Apr. 2020, doi: 10.1609/aaai.v34i05.6412.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[9] Q. Chen, Z. Zhuo, and W. Wang, "BERT for Joint Intent Classification and Slot Filling," Feb. 28, 2019, arXiv: arXiv:1902.10909. doi: 10.48550/arXiv.1902.10909.

[10] A. Inoue, M. Prado, and F. Cozman, "Automated Emergency Room Triage: Helping Patients Get the Best Treatment," in Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2020), Brasil: Sociedade Brasileira de Computação - SBC, Oct. 2020, pp. 591–602. doi: 10.5753/eniac.2020.12162.

[11] G. Gebrael et al., "Enhancing Triage Efficiency and Accuracy in Emergency Rooms for Patients with Metastatic Prostate Cancer: A Retrospective Analysis of Artificial Intelligence-Assisted Triage Using ChatGPT 4.0," Cancers, vol. 15, no. 14, Art. no. 14, Jan. 2023, doi: 10.3390/cancers15143717.

[12] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, "ParsBERT: Transformer-based Model for Persian Language Understanding," Neural Process. Lett., vol. 53, no. 6, pp. 3831–3847, Dec. 2021, doi: 10.1007/s11063-021-10528-4.

[13] Z. Rahimi and M. Shamsfard, "A Neuro Symbolic Approach for Contradiction Detection in Persian Text," JUCS - J. Univers. Comput. Sci., vol. 29, no. 3, pp. 242–264, Mar. 2023, doi: 10.3897/jucs.90646.

[14] A. Yoshida, Y. Kato, and S. Matsubara, "Revisiting Syntax-Based Approach in Negation Scope Resolution," in Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (*SEM 2023), A. Palmer and J. Camacho-collados, Eds., Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 18–23. doi: 10.18653/v1/2023.starsem-1.3.

[15] W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan, "A simple algorithm for identifying negated findings and diseases in discharge summaries," J. Biomed. Inform., vol. 34, no. 5, pp. 301–310, Oct. 2001, doi: 10.1006/jbin.2001.1029.

[16] H. Harkema, J. N. Dowling, T. Thornblade, and W. W. Chapman, “ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports,” J. Biomed. Inform., vol. 42, no. 5, pp. 839–851, Oct. 2009, doi: 10.1016/j.jbi.2009.05.002.

[17] M. Reyad, A. M. Sarhan, and M. Arafa, “A modified Adam algorithm for deep neural network optimization,” Neural Comput. Appl., vol. 35, no. 23, pp. 17095–17112, Aug. 2023, doi: 10.1007/s00521-023-08568-z.