

Article

Online Textual Symptomatic Assessment Chatbot Based on Q&A Weighted Scoring for Female Breast Cancer Prescreening

Jen-Hui Chen ^{1,2,3,4,*} , Obinna Agbodike ⁵ , Wen-Ling Kuo ⁶ , Lei Wang ⁷ , Chiao-Hua Huang ¹,
Yu-Shian Shen ¹ and Bing-Hong Chen ¹

¹ Department of Computer Science and Information Engineering, Chang Gung University, Taoyuan 33302, Taiwan; kuloyolu22@gmail.com (C.-H.H.); ivan880310@gmail.com (Y.-S.S.); bruce011043@gmail.com (B.-H.C.)

² Center for Artificial Intelligence in Medicine, Chang Gung Memorial Hospital, Taoyuan 33375, Taiwan

³ Artificial Intelligence Research Center, Chang Gung University, Taoyuan 33302, Taiwan

⁴ Department of Electronic Engineering, Ming Chi University of Technology, Taishan District, New Taipei City 24301, Taiwan

⁵ Division of Computer Science and Information Engineering, Department of Electrical Engineering, Chang Gung University, Taoyuan 33302, Taiwan; d0721009@cgu.edu.tw

⁶ Division of Breast Surgery and General Surgery, Department of Surgery, Chang Gung Memorial Hospital, Linkou and Taipei Branches, Taoyuan 33375, Taiwan; sylvie5285@gmail.com

⁷ School of Software, Dalian University of Technology, Dalian 116024, China; lei.wang@dlut.edu.cn

* Correspondence: jhchen@mail.cgu.edu.tw



Citation: Chen, J.-H.; Agbodike, O.; Kuo, W.-L.; Wang, L.; Huang, C.-H.; Shen, Y.-S.; Chen, B.-H. Online Textual Symptomatic Assessment Chatbot Based on Q&A Weighted Scoring for Female Breast Cancer Prescreening. *Appl. Sci.* **2021**, *11*, 5079. <https://doi.org/10.3390/app11115079>

Academic Editor: Carlos A. Iglesias

Received: 30 March 2021

Accepted: 28 May 2021

Published: 30 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The increasing number of female breast cancer (FBC) incidences in the East predominated by Chinese language speakers has generated concerns over women's medicare. To minimize the mortality rate associated with FBC in the region, governments and health experts are jointly encouraging women to undergo mammography screening at the earliest suspicion of FBC symptoms. However, studies show that a huge number of women affected by FBC tend to delay medical consultation at its early stage as a result of factors such as complacency due to unawareness of FBC symptoms, procrastination due to lifestyle, and the feeling of embarrassment in discussing private matters especially with medical personnel of the opposite gender. To address these issues, we propose a symptomatic assessment chatbot (SAC) based on artificial intelligence (AI) designed to prescreen women for FBC symptoms via a textual question-and-answer (Q&A) approach. The purpose of our chatbot is to assist women in engaging in communication regarding FBC symptoms, so as to subsequently initiate formal medical consultations for early FBC diagnosis and treatment. We implemented the SAC systematically with some of the latest natural language processing (NLP) techniques suitable for Chinese word segmentation (CWS) and trained the model with real-world FBC Q&A data obtained from a major hospital in Taiwan. The results from our experiments showed that the SAC achieved very high accuracy in FBC assessment scoring in comparison to FBC patients' screening benchmark scores obtained from doctors.

Keywords: female breast cancer (FBC); chatbot; patient-centric healthcare; NLP; trigram; CWS

1. Introduction

In recent years, the main reason for the high mortality rate associated with female breast cancer (FBC) cases as reported by several global health organizations is attributed to delayed diagnosis and late treatment. Several studies [1–3] have revealed that lack of awareness, hesitancy, and shyness to discuss FBC symptoms during its early stages are among the top contributing factors exposing women to the danger of late treatment. Previous research has shown that Asian women were considered the least vulnerable to FBC compared to women in other parts of the world [4], however, recently there is an unprecedented rapid rise in the rate of FBC incidences in Asia [3–5]. In Taiwan; for example, the government has deployed several medical vans across the country, known as

mobile mammography units [6], to serve as a desperate community outreach initiative to encourage screening for FBC symptoms. Unfortunately, irrespective of the costs of such an initiative, several patient-centric drawbacks concerning privacy, COVID-19 scare, lack of knowledge of FBC symptoms, etc., render such efforts less effective. Although the precise cause of FBC and its prevalence remains unascertained [7], the first step to averting these life-threatening risks requires that the affected persons address their symptoms in a timely manner by communicating with the appropriate healthcare providers.

In medical practice, before diagnosis and subsequent treatment are administered to patients, bilateral interactions based on question-and-answer (Q&A) is the basic and most vital approach used by doctors to fathom the type and stage of patients' internal symptoms (such as early-stage FBC), which are usually invisible to the eyes by mere observation. Therefore, effective communication between doctors and patients is generally advantageous towards achieving successful medicare outcomes [8], and telemedicine via chatbots is the promising modern way to foster it [9,10]. Based on this predication, various healthcare establishments are beginning to adopt interactive text-based chatbot systems [11] to streamline communication between doctors and their remotely distributed patients; with artificial intelligence (AI) technologies, chatbots are quickly gaining the potential to facilitate interactive communication-dependent services and the capability to independently make adaptive judgment calls reliably in task-oriented closed domains.

In this paper, we present the FBC symptomatic assessment chatbot (SAC) which is a Chinese-text-based Q&A interactive dialog system developed for Chang Gung Memorial Hospital (CGMH) in Taiwan. The aim of our research study is to proffer a solution to the patient-centric communication barriers hindering most women from initiating consultations for mammography screening, especially at the early stage of FBC symptoms. For this task, first, we understand that, in critical vertical domains such as the healthcare sector, it is important that a chatbot should be built to process and match the textual Q&A dialog contexts with high-level logical consistency. According to Chen et al. [12], the degree of dialog output accuracy is often the primary criteria used in judging the performance of a Chatbot.

However, with respect to achieving high output accuracy, modeling a healthcare chatbot in Chinese presents both technical and linguistic challenges, especially with regards to Chinese word segmentation (CWS) [13] and other factors pertaining to medical ethics that often limit access to substantial patients' medical data samples [11,14] required for the training of AI models. Nonetheless, we implemented the SAC model with a light-weight trigram-based natural language processing (NLP) algorithm trainable with low volume data. In addition, we created a web and an iOS front-end chat interface for easy access to the services of the chatbot. Figure 1 illustrates how the SAC achieves its goal of connecting prospective FBC patients to the healthcare providers over the network.

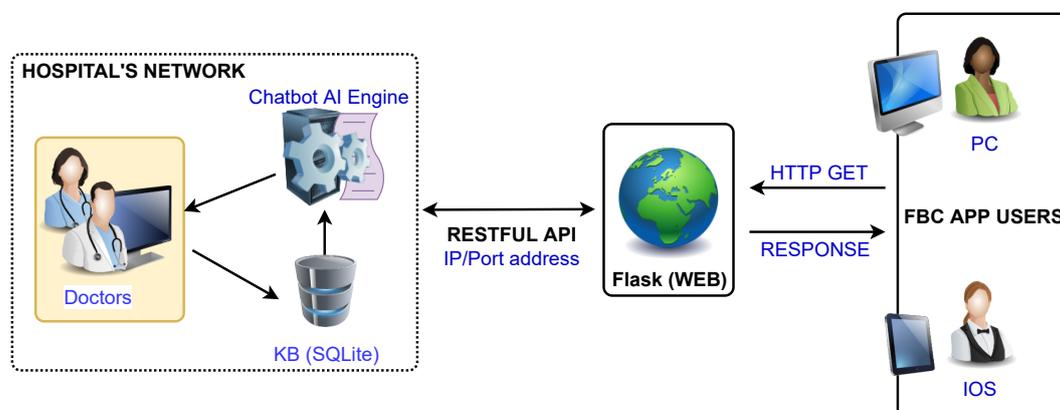


Figure 1. A simple network structure on how SAC connects prospective FBC patients to the healthcare provider.

Against this background, we affirm that our proposed SAC dialog system is a resourceful application capable of contributing immense benefits to both the end-users and

the healthcare domain responsible for handling FBC incidences. Such benefits include but are not limited to the following:

- Benefit to users: SAC's Q&A assessment raises the awareness of FBC ailment and encourages self examination of the breasts during dialog in the comfort of the users' valued privacy, which will stimulate them to communicate freely and sincerely about their FBC symptoms. In congruence to this regard, an earlier study by Lucas et al. [15] revealed that patients are more comfortable responding honestly to computer-administered health assessments than that of human physicians due to the "intimate" nature of information usually required in such conversational contexts. On the other hand, SAC has the ability to promptly schedule medical appointment for the users scored with positive FBC symptoms in-order to expedite timely FBC diagnosis and early medical treatment.
- Benefit to hospitals: SAC offers the potential benefit of controlling undue and inundating influx of patients in FBC healthcare departments, most of whom may schedule FBC medical appointments while having unrelated symptoms (i.e., FBC false negative cases). According to Bibault et al. [14], dialog systems can allow physicians to spend more time in treating patients who need the most consultation. On this account, the healthcare sector of many countries are beginning to utilize the digital services of chatbots for patients' disease prescreening assessment and triage in-order to reduce stress on health providers [16]. Therefore, by adopting SAC, hospitals can drastically minimize wastage of time, money, and other resource costs, and more importantly, doctors and other healthcare personnel will be enabled to work more productively to achieve effective patient-centric care service delivery.

The remainder of this paper is structured as follows: Section 2 covers a review of related works, and thereafter, in Section 3, we present the model description comprising of the adopted concepts used in implementing the SAC model. Section 4 contains brief details about the experiment and the corresponding result outcomes, respectively. Finally, Section 5 summarizes the work and provides a conclusion.

2. Related Works

Modern-day health service delivery is becoming more patient-centric than ever before. This new norm is instigating medical institutions across the world to seek urgent solutions on how to prioritize patients' medicare outcomes because a greater percentage of their reputation and profitability now depends on it [17]. Meanwhile, meeting every care need of patients is currently almost impossible for human physicians, considering the bottlenecks of gross maldistribution and shortage of healthcare workforce in the real-world scenario. Therefore, the paradigm and demand for patient-centric healthcare nowadays is introducing a new wave of physical and emotional stress problems for doctors and other health workers [18], which can result in low efficiency in delivering essential healthcare interventions and services.

On the basis of the above-stated problem, a plethora of diverse research studies has explored the use of text-based chatbots as a solution to support doctors and patients towards effective healthcare service delivery in various areas of medicine. To this regard, scholars such as Omoregbe et al. [19] developed an SMS-based NLP technique engineered with a knowledge base in the domain of known disease facts to assist medical experts in successfully boosting medical service delivery. Additionally, other similar studies have widely covered the aspects of using chatbots to support healthy lifestyles [9], mental health [11], pregnancy care [20,21], etc. [22]; however, very few related works have been conducted on the subject of oncology. With respect to breast cancer, Bibault et al. [14] developed an informative chatbot, referred to as Vik, which provides support to medically confirmed FBC patients and is capable of diagnosing other diseases as well. A subsequent study by Chaix et al. [23] investigated the effectiveness of the Vik chatbot from the patient-centric point of view and discovered that the use and acceptance of chatbots in healthcare can also improve patients' medication adherence. Judging from the perspective of Palan-

ica et al. [22], the integral generic features of text-based chatbots such as being gender neutral, racially unbiased, cost-effective, and available round-the-clock to ask or respond coherently to chats when needed are among the appealing qualities that makes chatbots suitable in supporting patients' care needs.

Recently, modelling of chatbots with deep neural network (DNN) algorithms has become the in-vogue approach widely used for the purpose of achieving intelligent dialog outcomes. Meanwhile, Najafabadi et al. [24] and Wang et al. [21] considered this approach as being very data-intensive. The former [24] indicated that the output accuracy of some of the models based on DNN approach depends on the hugeness of the data size used for model training, which, as a result, often necessitates demand for extensive computational resources that can only be afforded by a limited audience. In confirmation, Shaikhline and Khovanova [25] stated that DNNs trained with small dataset often exhibit unstable behavior in terms of their output performance. Considering that data availability for research in the closed domain of healthcare sector are often either scarce [14,25,26] due to strict medical ethics enacted to protect the privacy of patients' information or that the available data are unusable [27] poses some concerns regarding the application of DNNs in this regard. Although authors such as Kapočiūtė-Dzikiene [28] have conducted experiments on DNN-based chatbot (e.g., stacked bidirectional long short-term memory (BiLSTM)) and achieved interesting results after training with small-scale dataset for a closed-domain task, such approaches may not yet be fully accredited for implementation in critical service domains such as the healthcare sector, where dialog errors can result to fatal implications [12,14]. Meanwhile, contrary to DNNs, scholars [9,14,16,19] have evidently demonstrated the practical effectiveness of modelling medical chatbots with rule-based NLP frameworks that are based on traditional hand-crafted machine-learning approaches. This approach is often based on pattern-matching [9,10,14,19] with a knowledge base for domain-specific tasks and requires low-volume data and less computational resources for model training in contrast to the DNN approach.

Further research directions relevant to our work borders on the implementation of text-based natural language understanding for dialog systems. On this subject, English and Chinese texts are the most researched languages [28]. A study by Zhang et al. [29] highlighted the current social phenomenon of today's multilingualism of which the authors [29] proposed a two-stage Chinese–English mixed text normalization module to benefit NLP preprocessing tasks, owing to the recent prevalent norm of Chinese-speaking locals frequently mixing informal Chinese texts with English words, especially in social messaging platforms. Similarly, Shao et al. [30] have investigated greedy search algorithm with probabilistic N-gram matching, proposed for Chinese-to-English neural machine translation task. With precise regard to Chinese texts, several authors [31–33] unanimously agree that Chinese sentences being character sequences without word delimiters makes Chinese word segmentation (CWS) a key prerequisite in NLP. Ma and Hinrichs [31] proposed an embedding-based matching model for CWS using a greedy search segmenter to extract and match the historical and predicted Ngram-based character context features (i.e., unigrams and bigrams). In variance to [31], Cai et al. [32] implemented greedy search with a neural scorer (i.e., LSTM) to achieve fast and accurate CWS. Nevertheless, Fu et al. [33] conducted several experiments on eight diverse NLP models trained on seven different datasets, after which the authors proposed an attribute-aided evaluation method for CWS on the basis of their analytical research discovery that CWS evaluation performance varies per model and depends on the type of dataset used for training.

From the above reviewed studies, we generated the concept to model our proposed FBC symptomatic assessment chatbot (SAC) for best performance using the traditional handcrafted NLP approach. Our method also involved the application of trigram-based Chinese-text segmentation with probabilistic greedy search to effectively achieve contextual feature matching of the Q&A chat, in a domain-specific rule-based neural architecture. The justification for adopting this approach was necessitated by the limited size of the Q&A

medical data availed by our research sponsor (i.e., Chang Gung Memorial Hospital) for this work.

3. Methodology

Generally, the healthcare service industry is a critical rule-based sector with low tolerance for professional misconduct. With this understanding, we devised the SAC dialog system to be domain-specific, i.e., based on adaptive logical rules bound within the locus of FBC data in the system's knowledge base. In other words, the SAC only asks and responds to textual chats related to the subject of breast cancer symptoms and ailments. The main reason behind this mechanism is to keep the SAC task-oriented in order to prevent any tendencies of deviation towards out-of-scope conversational outcomes and ambiguities. Therefore, human emotional cues were deemed unnecessary in this study and, thus, were not considered.

As proof of our compliance with medical ethics and other best practices adopted in this research, the SAC was granted full approval for clinical trial by the institutional review board (IRB) of Chang Gung Medical Foundation with number: 202000815B0C501. In this section of the paper, we present the most important aspects of the SAC's concepts and modus operandi.

3.1. The Symptomatic Assessment Process

According to Fadhil and Schiavo [34], chatbots can gain self-learning ability through pattern matching to provide more natural interaction with users. Being a domain-specific rule-based neural architecture, the SAC is integrated with the SQLite database containing structured FBC medical Q&A data curated by the physicians of the general breast surgical department at the Linkou branch of CGMH, Taiwan. In compliance with medical ethics and best practices, we anonymized patients' information from the obtained data by filtering out personal details as was performed by Bibault et al. [14]. Thereafter, in conjunction with two experienced FBC specialist doctors in CGMH, we structured the symptomatic assessment knowledge-base questions and their corresponding answer probability matches.

3.1.1. Question and Answer Distribution

The data structure in SAC's knowledge base is formatted to aid the chatbot's logic adapter mechanism to effectively match Q&A's and to retrieve the appropriate next sub-question. In respect to this and together with the directive of the FBC specialist doctors, we adapted a total of 16 main question categories (QCs) unevenly distributed and hierarchically structured to cover every potential topics regarding female breast cancer ailment and symptoms (see Table 1 and the statistical representation in Figure 2). The agent can ask several sub-questions under each question category in Table 1 in a sequence-to-sequence manner (as illustrated in Figure 3). In other words, each response to SAC's query provided by the user is computed by an adaptive logical process module in the system's backend (i.e., the chatbot's engine), which utilizes the Q&A contextual match to determine the next relevant sub-question or next question category to retrieve from the knowledge base. This process summarizes the dialog flow handling process (as shown in Figure 3). Additionally, since the SAC asks the questions and utilizes the context of the users' response to determine next relevant main or sub-question to be retrieved from the knowledge base, the multi-turn response selection design problem that commonly poses a challenge in retrieval-based dialog systems is thus eliminated [12]. Moreover, according to Figure 4, each of the 16 main category of questions on FBC symptoms contains a varied distributed number of sub-questions (see Figure 2) to enable the SAC agent to understand the extent of the user's FBC symptoms as a human physician (i.e., doctors) would.

Table 1. The 16 major question categories (QC) structured in SAC’s knowledge base for chat assessment of users’ FBC symptoms. Positive user responses to SAC’s queries with * imply serious FBC symptoms, likewise, ‡ = moderate symptoms, while † = mild symptoms.

Question Topic Categories			
Index	Query	Index	Query
QC1 *	Did you feel any breast lumps?	QC9 ‡	Have you ever had a breast ultrasound examination?
QC2 †	Does the breast hurt?	QC10 ‡	Have you ever had a mammogram?
QC3 *	Is there any discharge from the nipple?	QC11 †	Have you ever had a breast tumor lab test or breast tumor removal surgery?
QC4 *	Is the appearance of the breast sunken, bulged, or wounded?	Q12 †	Have you ever had an underarm examination or breast augmentation surgery?
QC5 ‡	Is any of the breast skin itchy?	QC13 †	Have you ever had a breast MRI examination?
QC6 *	Do you personally have a medical history of ovarian cancer, breast cancer, prostate cancer, or pancreatic cancer?	QC14 *	Have you had any other imaging tests in the past that show a problem with the breast?
QC7 †	Do you have direct blood relatives or siblings suffering from metastatic prostate cancer or pancreatic cancer?	QC15 *	Have you ever had a blood or genetic test confirming you are in high-risk group for breast cancer?
QC8 †	Do you have a family history of breast cancer or ovarian cancer?	QC16 *	Have you or any of your family members tested positive for breast cancer gene (BRCA)?

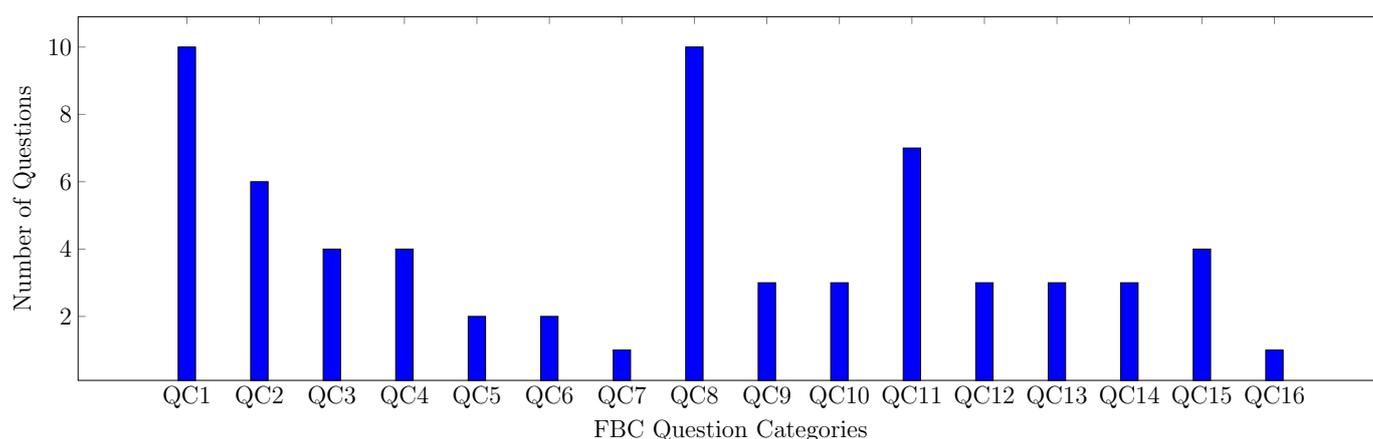


Figure 2. Statistical chart of questions’ topic categorial distribution.

3.1.2. Knowledge-Base Data Structure

The most frequent and notable symptoms of FBC ailment include but are not limited to users’ evident feeling of lumps, persistent feeling of pains, fluid discharge from the nipple, and visible changes on the skin area of the breasts, illustrated and denoted as s_1, \dots, s_4 in Figure 5. These symptoms often trigger human physician’s suspicion of a positive case of breast cancer before mammography screening is conducted. In the question category distribution (detailed in Table 1), most symptoms such as these were assigned high scores. Meanwhile, the FBC symptomatic assessment answer scores range on a scale of 1 to 10 (i.e., low to high). Take, for example, in the sample question 4 (QC4) in Figure 4, the agent queries the user if there is any visible depression, bulge, or lesion on the breast; therefore, assuming the user’s response correlates with the context of A1 in the answer match probability (i.e., s_4 symptom in Figure 5), the SAC’s logic module considers the user’s FBC symptoms as very serious (i.e., after matching the Q&A context with the domain lexical keywords in the knowledge base), and thus, the agent weighs and scores the user’s response as a 10 as predefined in the knowledge base (see Figure 4). Subsequently, such prospective FBC patients (scored high) is granted priority with respect to scheduling of doctor’s appointment at the end of the chat session.

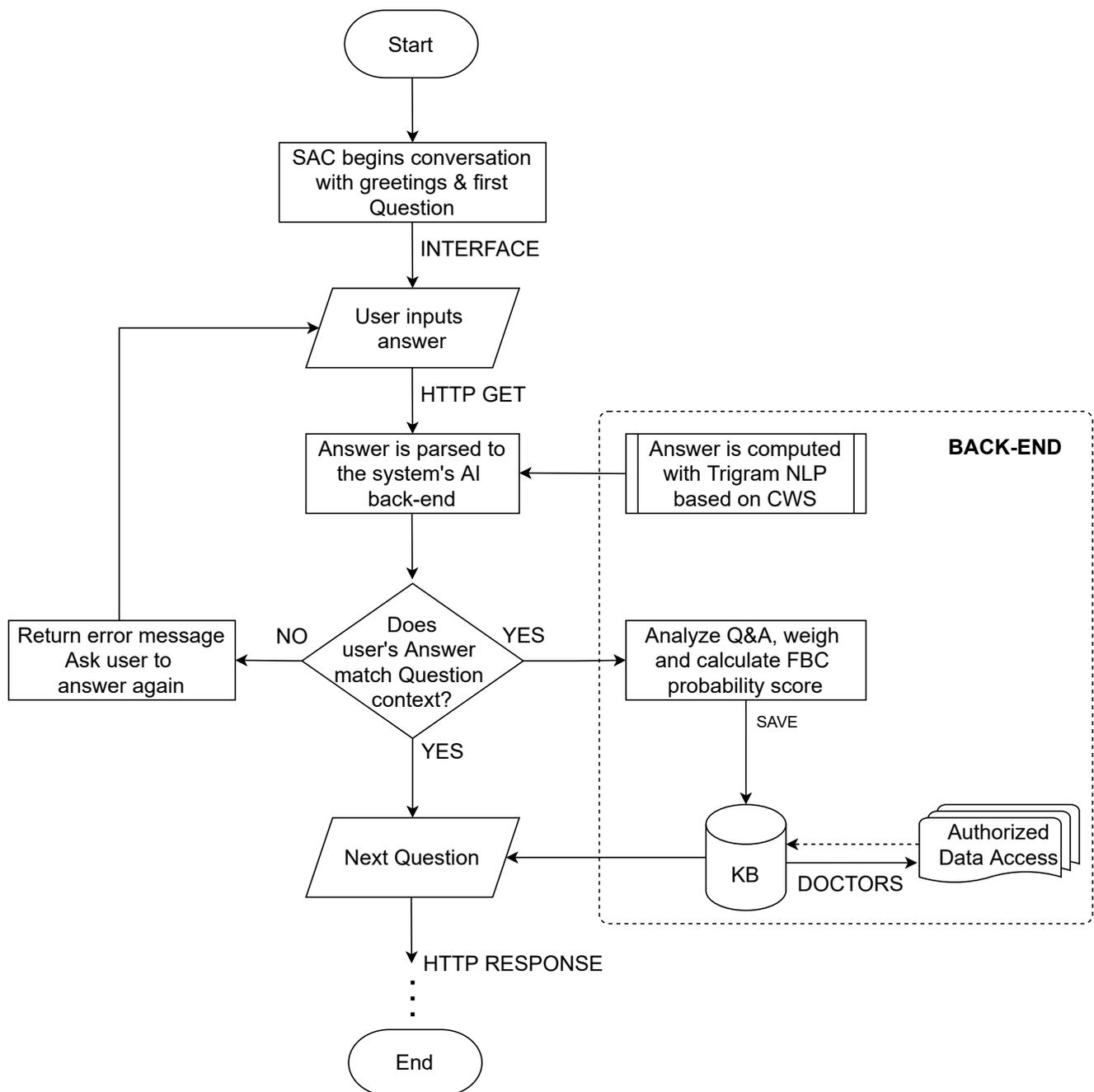


Figure 3. The dialog information flow process of the SAC model.

Index	Question Content	Index	Answer Match Probability	Score	Lexical Keywords
QC4*	乳房外觀是否凹陷、凸起或有傷口? Is the appearance of the breast sunken, bulged, or wounded?	A1	我左邊的胸部看起來有凹下去一塊 My left breast looks sunken.	10	左邊;凹下去 Left; sunken;
		A2	我右邊的胸部看起來有凹下去一塊 My right breast looks sunken.	10	右邊;凹下去 Right; sunken;
.....
		A16	不知道 I don't know. Not sure.	7	不知道; 不清楚 Don't know; Not sure
		A17	有 Yes	10	是的 Yes
			沒有 No	3	不 No

Index	Sub-Question Content	Index	Answer Match Probability	Score	Lexical Keywords
A1-Q1	有凹下去的狀況持續多久? How long have sunken condition persisted?	A1-Q1-A1	我有凹下去一塊的狀況大約 1 天 I have the sunken condition for about 1 day.	1	凹下去; 1 天 Sunken; 1 day
		A1-Q1-A2	我有凹下去一塊的狀況大約 1 個禮拜 I have a sunken condition for about 1 week.	1	凹下去; 1 週 Sunken; 1 week;
.....
		A17	我有凹下去一塊的狀況大約半年 I have a sunken condition for about half a year.	9	凹下去; 半年 Sunken; a half year; 6 months
		A18	我有凹下去一塊的狀況大約 3 年 I have a sunken condition for about 3 years.	9	凹下去; 3 年 Sunken; 3 years

Figure 4. Sample of the Q&A assessment knowledge-base data structure indicating how the context of a user’s response to a main question category (e.g., QC4) query is matched and scored by the SAC, which also proceeds to query the user further with a sub-question regarding the symptom.

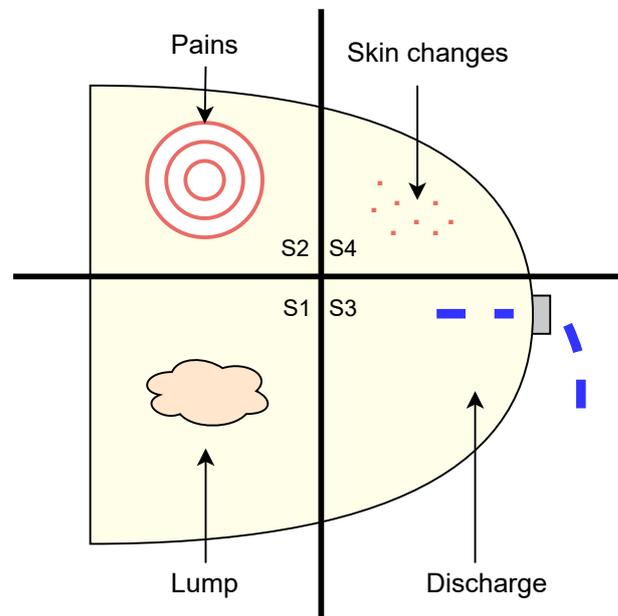


Figure 5. Schematic cross section of a breast showing the major stages of FBC ailment (s1, . . . , s4). The SAC is modeled to sequentially compute and score the seriousness of the FBC symptoms (if any), during textual Q&A dialog with the user.

3.1.3. Weighing and Scoring Module

From our several liaisons with the FBC specialists doctors, we understood how different levels of FBC symptoms are classified during human screening assessment of FBC patients. Based on that, we classified the FBC symptomatic assessment questions, i.e., the

16 main QCs and their sub-questions (Q), into three levels. Let A , B , and C denote the sets of the levels A , B , and C , respectively, and N_A , N_B , and N_C be the total number of questions that the patients answer at levels A , B , and C . We hereby note that each main question is accompanied by a varying number of sub-questions depending on the chat response answer provided by the subject (i.e., the patient/user). Let $Q_{i,j}^A$, $Q_{i,j}^B$, and $Q_{i,j}^C$ denote the value of main question i accompanied by sub-question j , which the subject answered during their chat with SAC. For example, $Q_{3,0}^A$ stands for the value of the subject's response answer to main question 3, which is assigned to A (i.e., $Q_{i,j}^A \in A$). In other words, $Q_{3,2}^A$ stands for the value of the response by the subject for main question 3 accompanied by a second sub-question. For the FBC symptom score evaluation, we assigned three weight factors of different values denoted as f_A , f_B , and f_C to each of the three class levels allotted to each of the questions. The values of f_A , f_B , and f_C were adjudged and provided by the CGMH FBC specialist doctors. Meanwhile, the summation of f_A , f_B , and f_C is 1 (i.e., $f_A + f_B + f_C = 1$). Based on this rule, the SAC computes the final FBC assessment score, denoted as S_1 , with the derived algorithm given below:

$$S_1 = \alpha \left(\frac{f_A}{N_A} \sum_{i=1}^{16} \sum_{j=0}^n Q_{i,j}^A + \frac{f_B}{N_B} \sum_{i=1}^{16} \sum_{j=0}^n Q_{i,j}^B + \frac{f_C}{N_C} \sum_{i=1}^{16} \sum_{j=0}^n Q_{i,j}^C \right), \quad (1)$$

where n is the number of sub-questions under a main FBC question topic, $Q_{i,j}^A = [1, 10]$, $Q_{i,j}^B = [1, 10]$, $Q_{i,j}^C = [1, 10]$, and α is an adjustment coefficient. We note that the coefficient α is an important variable in Equation (1) for accurate scaling of the SAC's assessment scoring the users for FBC symptoms. Although users' positive answers to level A questions for example are considered and scored by SAC as very high probability of being at risk of having FBC symptoms, further confirmation is needed, such as a clinical evaluation by a human FBC specialist doctor. The experimental setup values of the weight factors f_A , f_B , and f_C and the coefficient α are further discussed in Section 4.

3.2. Natural Language Understanding

The effectiveness of chatbots in the healthcare domain requires advanced reasoning capability to formalize medical knowledge [19], and this process begins with understanding the language vocabulary. Bearing in mind that Chinese sentences are strings of characters with no natural white spaces [31–33] unlike English texts and other alphabet-based language words, it is imperative to use CWS to disambiguate Chinese texts for effective natural language understanding (NLU). Despite the approach used, past evidences in existing literature acknowledged CWS as a sequence labeling problem in which the target Chinese word character influences the feature prediction [31]. Similar to Zhao et al. [35], we first used a greedy-search character tokenizer to represent the Chinese texts as vector embedding, then we fed the vectors as the input of a trigram module. Thereafter, the NLU output was parsed into the multi-logic adapters in the chatbot's engine for Q&A matching.

Since our task is domain specific, we adapted the method by Zeng et al. [36] and constructed the trigram feature representation function, expressed as follows:

$$\rho(w_1, w_2, \dots, w_\ell) = \prod_{i=1}^{\ell} \rho \left(\frac{w_i}{w_{i-(\tau-1)} w_{i-(\tau-1)+1} \dots w_{i-1}} \right), \quad (2)$$

where ℓ is the word character length, τ is the length of the trigram to be considered, and ρ is the probability of an i th word character w_i position corresponding to a learned j th word w_j as long as w_j is in the agent's knowledge base. This process is illustrated in the model architecture in Figure 6.

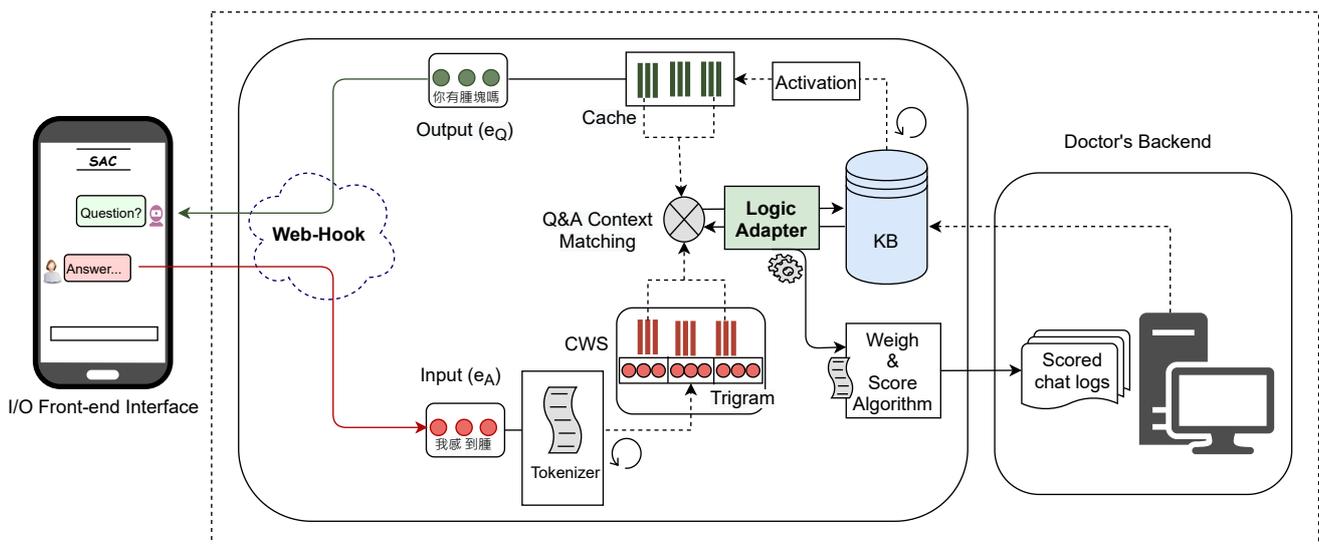


Figure 6. An overview of the SAC dialog system architecture. Herein, e_A represents the user's input answer embedding, whereas e_Q represents the agent's output question embedding.

3.3. The Logic Reasoning Module

To enrich the SAC with NLP and machine learning capabilities, we employed the ChatterBot (<https://chatterbot.readthedocs.io/en/stable/logic/index.html>, accessed on 10 February 2021) framework as the AI engine. The ChatterBot platform is a stable open-source python-based framework that comprises scalable multiple logic adapters, which we adopted in SAC to serve as the reasoning module (see Figure 6). We aligned the logic adapters to compute Equation (1) and to match the FBC symptomatic assessment Q&A during an ongoing chat session with a user. In addition, this also enables the agent to retrieve the most appropriate and contextual next response (i.e., question) that has the highest confidence score (also referred to as *the best match*) from the SAC's knowledge base.

4. Experimental Results

We implemented the SAC program with Python 3. The knowledge-base training data and the chat score log data are both configured in YAML (yet another multicolumn layout) format. Moreover, YAML is a data-serialization language readable by both humans and machines, which makes it generically suitable for our task. For the trigram module's CWS processing, we set the text encoding to UTF-8 to accommodate Chinese texts. In our approach, we used the double blind assessment method to verify the correctness of the proposed SAC method. The weight factors f_A , f_B , and f_C for SAC's FBC assessment score computation were originally assigned the values 0.6, 0.25, and 0.15, respectively, according to the directives of the doctors. Later on, the values of f_A , f_B , and f_C were slightly adjusted after we received the doctors' assessment score of each user. Other model hyper-parameter settings for the training were as default in ChatterBot python library. Due to data size limitation, we trained the SAC with the entire FBC Q&A medical data availed by the research sponsor CGMH.

Inasmuch as the performance outcome of text-based dialog models are popularly evaluated with a variety of computer-based metrics such as in [37], etc., we however subjected the performance of the SAC to expert human evaluators. According to Palanica et al. [22], human physicians are the traditional benchmark for treating patients over many centuries. Thus, their keen evaluation and approval of a medical chatbot's performance is crucial. For this reason, the SAC model's scoring mechanism (S_1) was fine-tuned to meet the doctors' scoring standard (denoted as S_2). The doctors used 135 anonymized patients' FBC Q&A screening data to evaluate the symptomatic assessment scoring capabilities of the SAC. The subsequent approval of SAC for clinical trial by IRB of Chang Gung Medical Foundation reflects the satisfaction of the evaluators with SAC's high accuracy and stable performance outcome.

4.1. Discussion

In the plot scenarios depicted in Figure 7, the red dots represents the users (i.e., FBC patients) and the values on the y-axis indicate the SAC’s FBC assessment scores, whereas on the x-axis is the doctor’s FBC screening scores which we considered in this study as the fixed benchmark standard. In case 1(a), it was observed that the performance of the SAC agent was very poor. Herein, for example, the batch of users that was scored 10 by the doctors received between 4.1 to 5.9 scores from the chatbot. Therefore, we introduced an adjustable scaling coefficient α in the algorithm in Equation (1) to address the low score outcome. Thereafter, in subsequent retraining of the SAC model, we experimentally varied the value of the α and observed improved scoring abilities of the agent as shown in Figure 7b,c, respectively. In Figure 7d, we adjusted the scaling value of α until we arrived at 1.53, and then also, we slightly modified the values of the weight factors to $f_A = 0.63$, $f_B = 0.22$, and $f_C = 0.15$, upon which, after completion of the training iterations, we achieved up to 95.4% assessment scoring accuracy. In this instance, only a few variations was observed between the agent’s score (S_1) and the FBC doctors’ score (S_2). For example, the lower end user’s FBC symptom that was scored 10 by the doctor was also likewise scored 9.2 by the chatbot, both of which are equally within the bounds of serious FBC ailment confirmation.

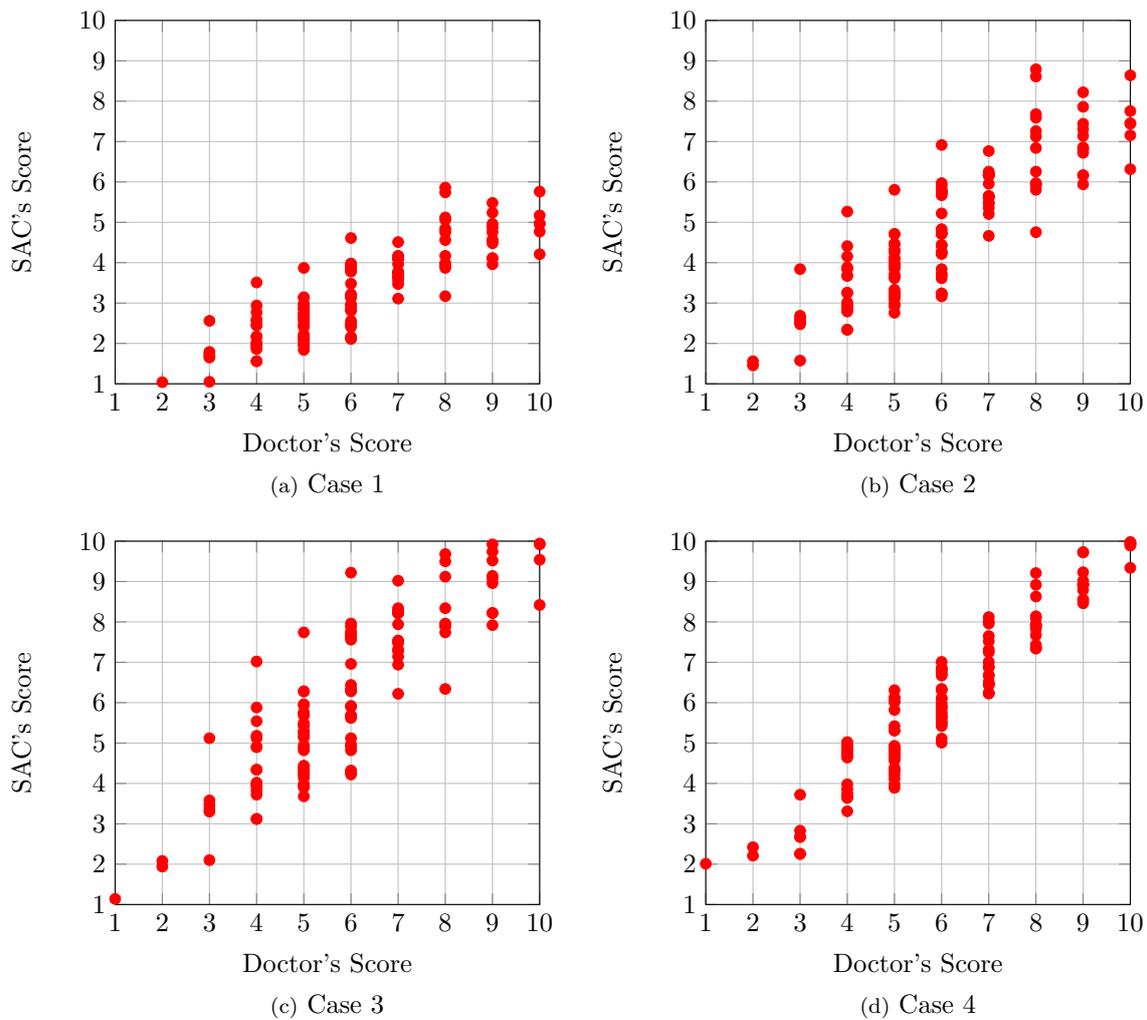
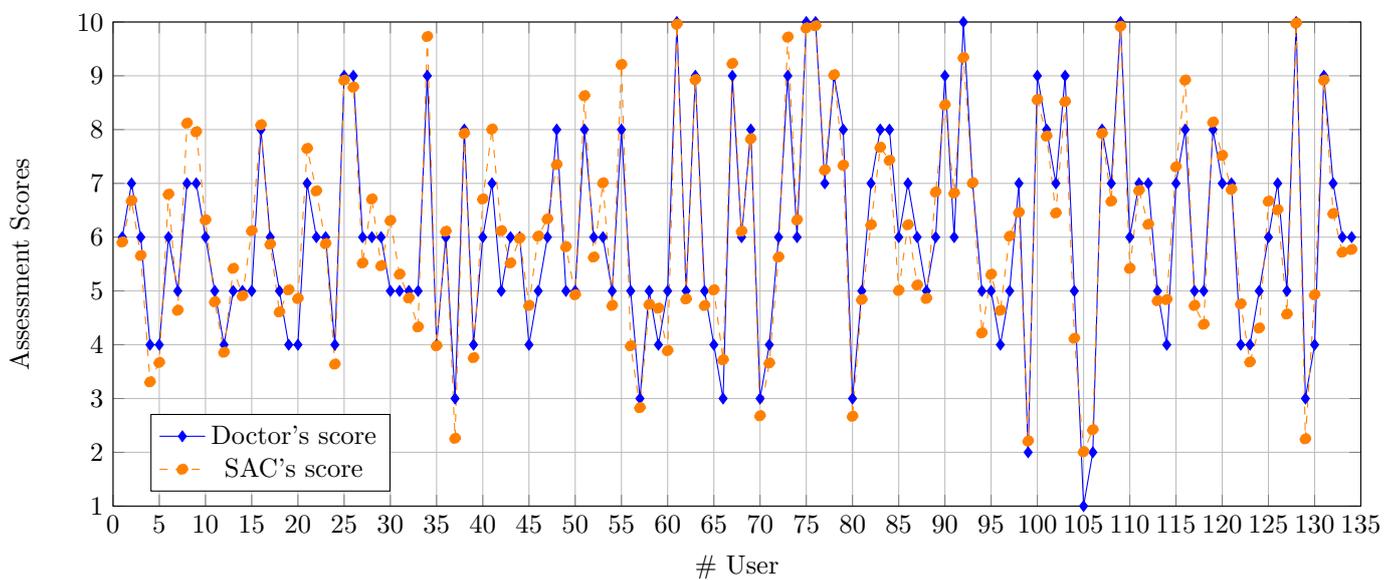
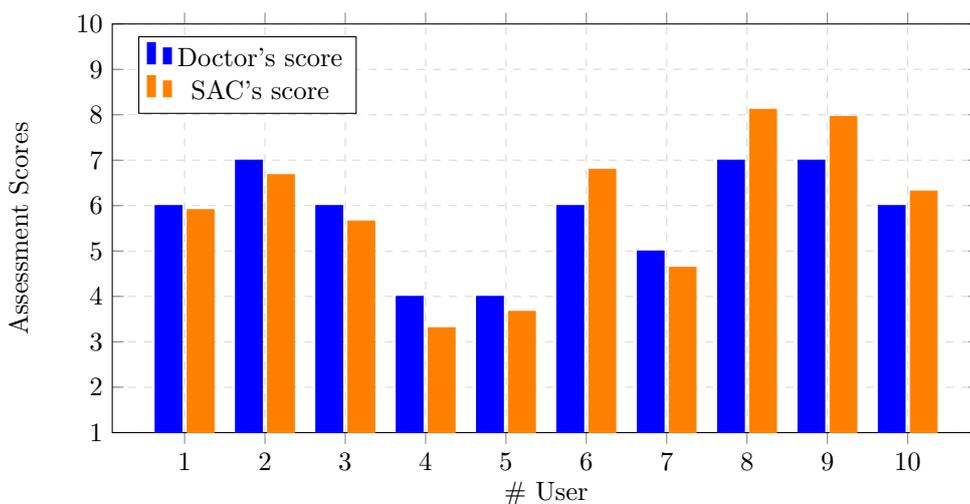


Figure 7. The comparison of statistical plots of SAC’s score vs. doctor’s score under different α , f_A , f_B , and f_C ; (a) $\alpha = 1$, $f_A = 0.6$, $f_B = 0.25$, and $f_C = 0.15$; (b) $\alpha = 1.45$, $f_A = 0.6$, $f_B = 0.25$, and $f_C = 0.15$; (c) $\alpha = 1.68$, $f_A = 0.6$, $f_B = 0.25$, and $f_C = 0.15$; and (d) $\alpha = 1.53$, $f_A = 0.63$, $f_B = 0.22$, and $f_C = 0.15$.

In Figure 8, we utilized more statistical representation approaches to further analyze SAC’s assessment scoring performance in comparison to the FBC assessment benchmark standard set by the FBC doctors. The FBC prescreening scores of SAC was compared against that of the doctors using the same group of patients. Since several distributed users can access and chat with SAC at the same time over the network, we subjected the SAC to multiple instances of textual dialogs using the same responses from the patients’ Q&A data. In the SAC’s testing process, a total number of 135 anonymized patients’ Q&A responses to the doctors’ FBC screening questions were used. In the graph plot of Figure 8a, the blue plot represents the doctors’ screening scores for each of the 135 patients, whereas the corresponding orange plot represents the SAC’s FBC assessment scores of the same patients. Keen observation of the graph plot outcome in Figure 8 indicates a high correlation accuracy achieved by SAC in FBC assessment scoring versus the benchmark scores from the CGMH FBC specialist doctors.



(a) The graph plot of all evaluation cases



(b) The histogram of the 1st evaluation case to the 10th evaluation case

Figure 8. Graphical plot and Histogram showing accuracy comparison between SAC’s assessment scoring of FBC patients versus the doctors’ benchmark FBC screening scores.

In addition, we also used the histogram in Figure 8b to show a cleaner comparative outcome of SAC's score accuracy versus the FBC doctors' scores using a random sample of 10 users. Here, the slight variations in both scores are clearly visible; however, irrespective of that, the histogram chart also shows that the SAC's scores meets up within the bounds of the doctors' score standard as described in Figure 7d.

4.2. Limitations

Although we designed the SAC's dialog-flow system to encourage users to respond to the agent's questions with domain-knowledge-related answers (see Figure 3), some users who tend to persist with verbose explanations in chats may result in an out-of-vocabulary situations for SAC as it is a rule-based chatbot. We assume that the likelihood of this occurrence can slightly affect SAC's overall assessment scoring. Therefore, in future work, we aim to increase the domain knowledge (i.e., FBC lexical keywords) in SAC's knowledge base and to improve the algorithm to become more dynamic in order to address these impending concerns.

5. Conclusions

In this paper, we presented the text-based chatbot (SAC) developed to prescreen users for symptoms of FBC ailment plaguing the majority of women around the world nowadays. The key design goal of SAC is to assist patients scored as having positive FBC symptoms to overcome personal barriers that often hinder them from initiating timely formal consultation with the appropriate healthcare providers. Invariably, this task is also a viable attempt to empower prospective FBC patients to avert the risk of uncertainties associated with late FBC treatment. Therefore, the SAC promises immense benefit to FBC healthcare providers and the general female population. To achieve the goal of this study, we modelled the SAC with bespoke trigram-based NLP technologies suitable for CWS. Finally, the evaluation outcome of SAC's performance indicate that it achieves high S_1 accuracy that is significantly comparable to the S_2 benchmark screening standard obtained from the FBC specialist doctors, which thus qualifies SAC for real-world practical use.

Author Contributions: Conceptualization, J.-H.C. and W.-L.K.; methodology, J.-H.C.; software, C.-H.H.; validation, J.-H.C., C.-H.H., and O.A.; formal analysis, J.-H.C., W.-L.K., and L.W.; investigation, J.-H.C.; resources, W.-L.K.; data curation, W.-L.K., C.-H.H., B.-H.C., and Y.-S.S.; writing—original draft preparation, O.A.; writing—review and editing, O.A. and J.-H.C.; visualization, O.A. and Y.-S.S.; supervision, J.-H.C. and L.W.; project administration, J.-H.C.; funding acquisition, J.-H.C. and W.-L.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 108-2221-E-182-042-MY2; and in part by the Chang Gung Memorial Hospital, Kweishan, Taoyuan, Taiwan, under Grants CMRPD2J0012 and CMRPD2I0052.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (IRB) of Chang Gung Medical Foundation (protocol code 202000815B0C501 and date of approval 27 May 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interests.

References

1. Ayaz, F.; Ayaz, S.B.; Farrukh, M. Reasons for Delayed Presentation of Women with Breast Cancer. *J. Islamabad Med. Dent. Coll.* **2016**, *5*, 187–191.
2. Tesfaw, A.; Demis, S.; Munye, T.; Ashuro, Z. Patient Delay and Contributing Factors Among Breast Cancer Patients at Two Cancer Referral Centres in Ethiopia: A Cross-Sectional Study. *J. Multidiscip. Healthc.* **2020**, *13*, 1391–1401. [[CrossRef](#)] [[PubMed](#)]
3. Zhang, H.; Wang, G.; Zhang, J.; Lu, Y.; Jiang, X. Patient delay and associated factors among Chinese women with breast cancer: A Cross-Sectional Study. *J. Med.* **2019**, *98*, e17454. [[CrossRef](#)]

4. Anderson, B.; Aranda, S.; Chatamra, K.; Cheung, P.; Chiou, S.-T.; Crossing, S.; Dent, R.; Fan, Z.; Ginsburg, O.; Han, S.; et al. *Breast Cancer in Asia: The Challenge and Response*; The Economist Intelligence Unit: London, UK, 2016.
5. Bhatia, H. Breast Cancer in Asia. General Reinsurance AG. November 2016. Available online: <https://media.genre.com/documents/ri16-4-en.pdf> (accessed on 28 January 2021).
6. Tsai, H.Y.; Chang, Y.L.; Shen, C.T.; Chung, W.S.; Tsai, H.J.; Chen, F.M. Effects of the COVID-19 pandemic on breast cancer screening in Taiwan. *Breast* **2020**, *54*, 52–55. [[CrossRef](#)] [[PubMed](#)]
7. Ataollahi, M.R.; Sharifi, J.; Paknahad, M.R.; Paknahad, A. Breast cancer and associated factors: A review. *J. Med. Life* **2015**, *8*, 6–11.
8. Moser, E.C.; Narayan, G. Improving breast cancer care coordination and symptom management by using AI driven predictive toolkits. *Breast* **2019**, *50*, 25–29. [[CrossRef](#)] [[PubMed](#)]
9. Fadhil, A.; Gabrielli, S. Addressing Challenges in Promoting Healthy Lifestyles: The AI-Chatbot Approach. In Proceedings of the 11th EAI international conference on pervasive computing technologies for healthcare, Barcelona, Spain, 23–26 May 2017; pp. 261–265.
10. Siddique, S.; Chow, J.C.L. Machine Learning in Healthcare Communication. *Encyclopedia* **2021**, *1*, 220–239. [[CrossRef](#)]
11. Hoermann, S.; McCabe, K.L.; Milne, D.N.; Calvo, R.A. Application of Synchronous Text-based Dialogue Systems in Mental Health Interventions: Systematic Review. *J. Med. Internet Res.* **2017**, *19*, e267. [[CrossRef](#)]
12. Chen, J.; Agbodike, O.; Wang, L. Memory-Based Deep Neural Attention (mDNA) for Cognitive Multi-Turn Response Retrieval in Task-Oriented Chatbots. *Appl. Sci.* **2020**, *10*, 5819. [[CrossRef](#)]
13. Chang, K.-C.; Chang, H.-T. Is It Possible to Use Chatbot for the Chinese Word Segmentation? In Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval, Tokushima, Japan, 28–30 June 2019; pp. 20–24.
14. Bibault, J.E.; Chaix, B.; Guillemassé, A.; Cousin, S.; Escande, A.; Perrin, M.; Pienkowski, A.; Delamon, G.; Nectoux, P.; Brouard, B. A Chatbot Versus Physicians to Provide Information for Patients With Breast Cancer: Blind, Randomized Controlled Noninferiority Trial. *J. Med. Internet Res.* **2019**, *21*, e15787. [[CrossRef](#)] [[PubMed](#)]
15. Lucas, G.M.; Gratch, J.; King, A.; Morency, L. It's only a computer: Virtual humans increase willingness to disclose. *Comput. Hum. Behav.* **2014**, *37*, 94–100. [[CrossRef](#)]
16. Martin, A.; Nateqi, J.; Gruarin, S.; Munsch, N.; Abdarrahmane, I.; Zobel, M.; Knapp, B. An artificial intelligence-based first-line defence against COVID-19: Digitally screening citizens for risks via a chatbot. *Nat. Sci. Rep.* **2020**, *10*, 19012. [[CrossRef](#)]
17. Majid, U.; Wasim, A. Patient-centric culture and implications for patient engagement during the COVID-19 pandemic. *Patient Exp. J.* **2020**, *7*, 5–16. [[CrossRef](#)]
18. Koinis, A.; Giannou, V.; Drantaki, V.; Angelaina, S.; Stratou, E.; Saridi, M. The Impact of Healthcare Workers Job Environment on their Mental-emotional Health. Coping Strategies: The Case of a Local General Hospital. *Health Psychol. Res.* **2015**, *3*, 12–17. [[CrossRef](#)] [[PubMed](#)]
19. Omoregbe, N.A.I.; Ndaman, I.O.; Misra, S.; Abayomi-Alli, O.O.; Damasevicius, R. Text Messaging-Based Medical Diagnosis Using Natural Language Processing and Fuzzy Logic. *J. Healthc. Eng.* **2020**, *2020*, 8839524. [[CrossRef](#)]
20. Vaira, L.; Bochicchio, M.A.; Conte, M.; Casaluci, F.M.; Guillemasse, A.; Melpignano, A. MamaBot: A System based on ML and NLP for supporting Women and Families during Pregnancy. In Proceedings of the 22nd International Database Engineering & Applications Symposium 2018, Villa San Giovanni, Italy, 18–20 June 2018; pp. 273–277.
21. Wang, L.; Wang, D.; Tian, F.; Peng, Z.; Zhang, X.; Ma, S.; Yu, M.; Ma, X.; Wang, H. CASS: Towards Building a Social-Support Chatbot for Online Health Community. *ACM Hum.-Comput. Interact.* **2021**, *5*, 1–31.
22. Palanica, A.; Flaschner, P.; Thommandram, A.; Li, M.; Fossat, Y. Physicians' Perception of Chatbots in Health Care: Cross-Sectional Web-Based Survey. *J. Med. Internet Res.* **2019**, *21*, e12887. [[CrossRef](#)] [[PubMed](#)]
23. Chaix, B.; Bibault, J.E.; Pienkowski, A.; Delamon, G.; Guillemasse, A.; Nectoux, P.; Brouard, B. When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients with Breast Cancer and a Chatbot. *JMIR Cancer* **2019**, *5*, e12856. [[CrossRef](#)] [[PubMed](#)]
24. Najafabadi, M.M.; Villanustre, F.; Khoshgoftaar, T.M.; Seliya, N.; Wald, R.; Muharemagic, E. Deep learning applications and challenges in big data analytics. *J. Big Data* **2015**, *2*, 1. [[CrossRef](#)]
25. Shaikhina, T.; Khovanova, N.A. Handling limited datasets with neural networks in medical applications: A small-data approach. *AI Med.* **2017**, *75*, 51–63. [[CrossRef](#)] [[PubMed](#)]
26. Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: Review, opportunities and challenges. *J. Healthc. Eng.* **2017**, *19*, 1236–1246. [[CrossRef](#)] [[PubMed](#)]
27. Shaikhina, T.; Khovanova, N.A. A Review of Challenges and Opportunities in Machine Learning for Health. *AMIA Summits Transl. Sci. Proc.* **2020**, *2020*, 191–200.
28. Kapočiušė-Dzikiėnė, J. A Domain-Specific Generative Chatbot Trained from Little Data. *Appl. Sci.* **2020**, *10*, 2221. [[CrossRef](#)]
29. Zhang, Q.; Chen, H.; Huang, X. Chinese-English Mixed Text Normalization. In Proceedings of the 7th ACM international conference on Web search and data mining, New York, NY, USA, 24–28 February 2014; pp. 433–442.
30. Shao, C.; Feng, Y.; Chen, X. Greedy Search with Probabilistic N-gram Matching for Neural Machine Translation. In Proceedings of the ACL Conference Empirical Methods in Natural Language Processing, Brussels Belgium, 31 October–4 November 2018; pp. 4778–4784.

31. Ma, J.; Hinrichs, E. Accurate Linear-Time Chinese Word Segmentation via Embedding Matching. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, Beijing China, 26–31 July 2015; pp. 1733–1743.
32. Cai, D.; Zhao, H.; Zhang, Z.; Xin, Y.; Wu, Y.; Huang, F. Fast and Accurate Neural Word Segmentation for Chinese. *arXiv* **2017**, arXiv:1704.07047v1.
33. Fu, J.; Liu, P.; Zhang, Q.; Huang, X. Rethink CWS: Is Chinese Word Segmentation a Solved Task? *arXiv* **2020**, arXiv:2011.06858v2.
34. Fadhil, A.; Schiavo, G. Designing for Health Chatbots. *arXiv* **2019**, arXiv:1902.09022v1.
35. Zhao, J.; Liu, H.; Bao, Z.; Bai, X.; Li, S.; Lin, Z. N-gram Model for Chinese Grammatical Error Diagnosis. In Proceedings of the 4th Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA 2017), Taipei, Taiwan, 1 December 2017; pp. 39–44.
36. Zeng, D.; Wei, D.; Chau, M.; Wang, F. Domain-specific Chinese word segmentation using suffix tree and mutual information. *Inf. Syst. Front.* **2010**, *13*, 115–125. [[CrossRef](#)]
37. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.-J. Bleu: A Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; pp. 311–318.