

Article

Medical Instructed Real-Time Assistant for Patient with Glaucoma and Diabetic Conditions

Ubaid Ur Rehman ^{1,2}, Dong Jin Chang ³, Younhea Jung ³, Usman Akhtar ¹,
Muhammad Asif Razzaq ¹ and Sungyoung Lee ^{1,*}

¹ Department of Computer Science and Engineering, Kyung Hee University (Global Campus), 1732 Deogyong-daero, Giheung-gu, Yongin-si, Gyeonggi-do 17104, Korea; ubaid.rehman@khu.ac.kr (U.U.R.); usman@oslab.khu.ac.kr (U.A.); asif.razzaq@oslab.khu.ac.kr (M.A.R.)

² School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad 44000, Pakistan

³ Department of Ophthalmology, Catholic University of Korea Yeouido Saint Mary's Hospital, Seoul 07345, Korea; hpalways@catholic.ac.kr (D.J.C.); younhea@hotmail.com (Y.J.)

* Correspondence: sylee@oslab.khu.ac.kr; Tel.: +82-31-201-2514

Received: 21 February 2020; Accepted: 22 March 2020; Published: 25 March 2020



Abstract: Virtual assistants are involved in the daily activities of humans such as managing calendars, making appointments, and providing wake-up calls. They provide a conversational service to customers around-the-clock and make their daily life manageable. With this emerging trend, many well-known companies launched their own virtual assistants that manage the daily routine activities of customers. In the healthcare sector, virtual medical assistants also provide a list of relevant diseases linked to a specific symptom. Due to low accuracy and uncertainty, these generated recommendations are untrusted and may lead to hypochondriasis. In this study, we proposed a Medical Instructed Real-time Assistant (MIRA) that listens to the user's chief complaint and predicts a specific disease. Instead of informing about the medical condition, the user is referred to a nearby appropriate medical specialist. We designed an architecture for MIRA that considers the limitations of existing virtual medical assistants such as weak authentication, lack of understanding multiple intent statements about a specific medical condition, and uncertain diagnosis recommendations. To implement the designed architecture, we collected the chief complaints along with the dialogue corpora of real patients. Then, we manually validated these data under the supervision of medical specialists. We then used these data for natural language understanding, disease identification, and appropriate response generation. For the prototype version of MIRA, we considered the cases of glaucoma (eye disease) and diabetes (an autoimmune disease) only. The performance measure of MIRA was evaluated in terms of accuracy (89%), precision (90%), sensitivity (89.8%), specificity (94.9%), and F-measure (89.8%). The task completion was calculated using Cohen's Kappa ($k = 0.848$) that categorizes MIRA as 'Almost Perfect'. Furthermore, the voice-based authentication identifies the user effectively and prevent against masquerading attack. Simultaneously, the user experience shows relatively good results in all aspects based on the User Experience Questionnaire (UEQ) benchmark data. The experimental results show that MIRA efficiently predicts a disease based on chief complaints and supports the user in decision making.

Keywords: primary healthcare; chief complaint; virtual medical assistant; spoken natural language; disease diagnosis; medical specialist

1. Introduction

With the emerging trends of technology, virtual assistants help users complete their daily routine tasks efficiently. Most of the virtual assistants use artificial intelligence and provide personalized

assistance to the users in the form of managing calendars, controlling smart environments, navigation, making an appointment, providing wake-up calls, and many more things [1]. Many applications from different domains currently have their own built-in virtual assistants such as televisions [2], mobile devices [3], vehicles [4], and the Internet of things [5,6]. The virtual assistant is also known as a chatbot, dialogue manager, virtual agent, interactive assistant, or conversational agent. Many well-known companies including Apple (Siri), Google (Assistant), Samsung (Bixby) and Amazon (Alexa) introduced their own virtual assistants. These virtual assistants provide an interactive user interface (text, speech, or both) that have the ability to understand requests, handle complex tasks, and generate an appropriate response using the machine learning model [7].

In the healthcare sector, the adoption of machine learning facilitated diagnosis [8,9], treatment [10,11], and streamlining of administrative tasks [12]. With the popularity of virtual assistants, healthcare is also moving toward this technology. It prevents unnecessary visits to the doctor, which reduces the administrative burden, increases efficiency, and support clinical decisions. According to a survey conducted in [13], primary care physicians spent more time managing electronic medical records (EMRs) than engaging with patients. Therefore, several virtual medical assistants were introduced, such as Nuance [14], Suki [15], and Robin Healthcare [16], which automate the process of documenting clinical information using artificial intelligence and provide services to the healthcare provider [17–19]. Moreover, several virtual medical assistants provide trusted information based on an analysis of medical symptoms, which include MedWhat [20], Your.MD [21], and Sensely [22]. These provide personal healthcare assistance using medical knowledge on the web, and EMR. These virtual medical assistants show a list of relevant diseases that match the input symptoms.

Suppose, a statement ‘I have abdominal pain’ is linked with a list of conditions, which include bowel cancer, constipation, Crohn’s disease, and gluten intolerance. The recommendation is predicted based on one symptom and requires approval from the medical specialist [23]. Due to the low accuracy and uncertainty of the existing virtual medical assistants, the resulting list of conditions may lead to depression, anxiety, and hypochondriasis [24]. To the best of our knowledge, none of the existing virtual medical assistants in the natural language processing domain considered real-time disease diagnosis based on the user’s chief complaint, which has the utmost priority, stated in the patient’s own words, and is the main reason for the patient’s visit. It may be possible that more than one disease has the same kind of chief complaint, making it hard to identify a specific disease. Furthermore, every person has their own accent and way of explaining the chief complaint, so understanding this type of conversation is a challenge for the virtual medical assistant as well.

In this study, we considered the challenges faced by the existing virtual medical assistant and proposed a solution in terms of the Medical Instructed Real-time Assistant (MIRA). MIRA supports primary healthcare services and uses spoken natural language for interactive communication to achieve a high success rate on task completion [25]. Moreover, MIRA analyzes the user’s chief complaint and predicts a specific disease. Then, the users are referred to a nearby appropriate medical specialist based on the predicted disease. For the prototype version of MIRA, we used the chief complaint of glaucoma and diabetes based on the availability of collaborative medical specialists from the Yeouido Saint Mary’s Hospital, Republic of Korea.

The main contributions provided by this study are summarized as follows:

- We introduced the MIRA that identifies a disease based on user’s chief complaint, understands single and multiple intent statements about a specific medical condition, and generates an appropriate response.
- We added an identity and access manager, a session manager, and security event logging and monitoring to the MIRA architecture. These provide strong authentication, manage the conversational state, and monitor the system for anomalies, respectively.
- We created a dataset of 816 patient chief complaints that were manually validated under the supervision of medical specialists, and were classified into glaucoma, diabetes, and other labels under the broad category of diseases.

- We designed stock phrases from the recorded 816 dialogue corpora that contain 11,532 utterances. Each utterance was manually annotated for intent and context identification.
- We evaluated MIRA based on a performance measure (including accuracy, precision, sensitivity, specificity, and f-measure), task completion, security, and user experience.

The rest of this paper is organized as follows. The overview of literature related to virtual medical assistants is described in Section 2. Then, Section 3 provides a comprehensive description of the MIRA methodology including system architecture, digital brain, and a case study. Subsequently, the evaluation of MIRA is presented in Section 4. Finally, Section 5 summarizes the work proposed in this study.

2. Related Work

We performed a systematic search of existing literature from the well-known digital libraries such as IEEE, ScienceDirect, ACM, Springer, PubMed, and Scopus. Based on this study, we focused on a spoken dialogue-based system that supports healthcare services. Therefore, we excluded the literature that does not focus on healthcare services and uses text, click, or touch as an interactive medium. Moreover, the studies that considered the Wizard-of-Oz concept were also filtered out. Based on these criteria, we found 14 studies and classified them into Finite State Assistants (10 studies) and Frame-based Assistants (4 studies). A comprehensive description of each category is provided in the subsequent sections.

2.1. Finite State Assistants

The finite state assistant asks a series of relevant questions to make a decision. This type of assistant does not support personalized recommendations because it follows the same sequential steps for each user. Philip et al. designed an Embodied Conversational Agent (ECA) for sleep disorder patients that ask questions using the Epworth Sleepiness Scale and identify the somnolence patients [26]. Similarly, the mental disorder diagnostic system conducts an interview based on DSM-5 criteria and identifies patients with major depressive disorders [27]. Moreover, an ECA was proposed for autism spectrum disorder patients that use audiovisual features for teaching social communication skills [28]. The proposed system is also effective for those experiencing social complications. To reduce hospitalization of suicidal patients, an e-caring avatar was proposed in [29], which involves patients in self-care conversations and recommends relevant videos. To monitor chronic pain patients, Levin et al., proposed a Pain Monitoring Voice Diary that asks a sequence of questions and identifies the severity of pain accordingly [30]. Moreover, the virtual agent for monitoring diabetic patients was proposed in [31], which makes a phone call once a week to collect vitals. Similarly, the spoken dialogue-based diabetic monitoring system collects patient vitals and helps physicians provide recommendations remotely based on the recorded information [32]. Virtual human interviewers are becoming popular due to anonymity and rapport building that supports posttraumatic stress disorder patients. Lucus et al., proposed a virtual human interviewer, which conducts an interview with military service members involved in an intense situation and identifies the symptoms associated with their mental state [33]. A similar kind of virtual agent was proposed in [34], which interacts with the users and identifies their mental symptoms using mixed methods for triangulation of data. Moreover, a rule-based patient-centric application was proposed in [35], which provides medical coaching services.

2.2. Frame-Based Assistants

The frame-based assistant analyzes and extracts the content from the user's conversation, then fills in the existing template to generate an appropriate response. The generated response may be personalized depending upon the business logic and training model of the corresponding virtual assistant. Ireland et al., proposed 'Harlie', which converses with the user on a variety of topics and helps in the neurological conditions of Parkinson's patients [36]. Similarly, a virtual nurse was

proposed in [37] to support maternal healthcare and provide guidance to expectant mothers during pregnancy. A few smartphone applications are also available that provide medical information after the analysis of symptoms such as MedWhat [20], Your.MD [21], and Sensely [22]. Giorgino et al., proposed a virtual medical assistant that interacts with hypertensive patients and collects relevant data, which help the physician to evaluate the risk of cardiovascular disease [38]. In [39], the virtual medical assistant supports general practitioners by analyzing patient health conditions (using a breast cancer ontological model) and recommending an oncologist.

2.3. Limitations of Existing Studies

According to our survey analysis, we identified three limitations in the existing studies that focused on spoken dialogue-based virtual medical assistants.

- None of the existing studies considered security as a primary factor except [30], which uses the traditional PIN-based authentication mechanism [40], and it is vulnerable to brute-force attack [41]. The virtual medical assistant interacts with users and gathers health-related information. The leakage of such information may lead to different attacks such as masquerading, and ransomware [42,43]. Moreover, commercially available applications such as Your.MD [21], and Sensely [22] only comply with the security standards.
- Most of the existing studies along with commercially available virtual medical assistants analyze the input symptoms, and either provide a list of specific diseases or relevant information [44]. None of the existing spoken dialogue-based system considered patient chief complaint corpora for disease prediction or medical advice.
- Limited studies focused on frame-based assistants due to various challenges such as intent identification, context awareness, and appropriate response generation. However, it provides interactions in a natural way (i.e., similar to humans) and keeps the user motivated to continue the conversation [45].

2.4. Medical Awareness Survey

We conducted a survey to assess medical awareness among university students and determine the need for MIRA. For this purpose, we designed a questionnaire and obtained approval from the Kyung Hee University Ethics Assessment Committee (KHU-EAC) after rigorous analysis of privacy aspects. The questionnaire was distributed via email among different departments including Computer Science and Engineering, Electrical and Electronic Engineering, Biomedical Engineering, Life Sciences, and Foreign Languages. The survey form was active for five consecutive working days. We received 119 responses from the age group (18 to 36 years) across 11 countries (International Students). Figure 1 presents the country-based distribution of participants along with gender ratio of male (50.8%) and female (49.2%). The participants responded to five polar questions as shown in Table 1. The survey result showed that 25% of the respondents had an awareness of medication and take medicine without doctor consultation (such as aspirin for pain and fever, amoxicillin for infection, and many more). These participants are also able to identify appropriate medical specialists based on their symptoms. The remaining 75% discuss with friends, family or general physicians. Healthcare services are expensive in most countries. Therefore, the majority of respondents preferred to discuss their symptoms with friends or family, which helps them to determine whether to seek an appropriate medical specialist. However, a small number of participants are not open to these discussions due to personal reasons. Overall, the majority of participants were excited about an application that understands speech-based natural language, determines specific disease based on chief complaints, and recommends a nearby appropriate medical specialist.

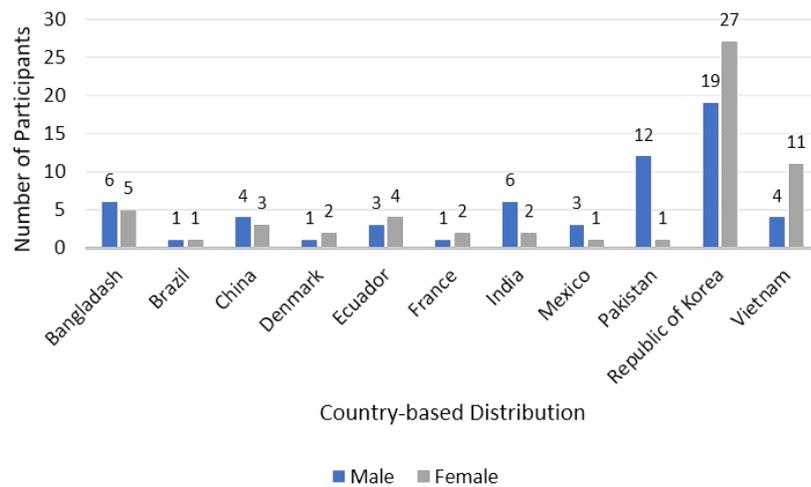


Figure 1. Medical awareness survey: Country-based distribution of participants along with gender ratio.

Table 1. Medical awareness survey questionnaire results.

Serial No.	Questions	Responses	
1	Do you have awareness of medication?	Yes (25%)	No (75%)
2	Are healthcare services expensive in your country?	Yes (83.3%)	No (16.7%)
3a	Based on your chief complaint, can you make a decision about an appropriate medical specialist?	Yes (25%)	No (75%)
3b	If you selected 'No' in 3a, then with whom will you discuss the situation?	Friends or Family (61.1%)	General Physician (38.9%)
3c	In the case of 'Friends or Family' in 3b, does the discussion help you to decide about the appropriate medical specialist?	Yes (63.6%)	No (36.4%)
4	Are you interested in a smartphone application that listens to your chief complaint and recommends a nearby appropriate medical specialist?	Yes (91.7%)	No (8.3%)
5	What type of interactive communication medium would you prefer for the smartphone application?	Speech-based (70.8%)	Text-based (29.2%)

3. Methodology

In this section, we deliver a comprehensive description of our designed state-of-the-art virtual medical assistant (MIRA), which provides efficient and reliable service to the user. First, we describe the overall system architecture of MIRA as shown in Figure 2, where the three modules (such as identity and access manager, session manager, and security event logging and monitoring) are introduced and integrated with the basic architecture (i.e., voice user interface, speech recognition, natural language understanding, and dialogue manager). Then, the next sub-section provides details about

the composition of the MIRA’s digital brain, which includes the knowledge source and stock phrases that support natural language understanding and appropriate response generation. Finally, we provide a case study at the end of this section that gives a better understanding of the MIRA.

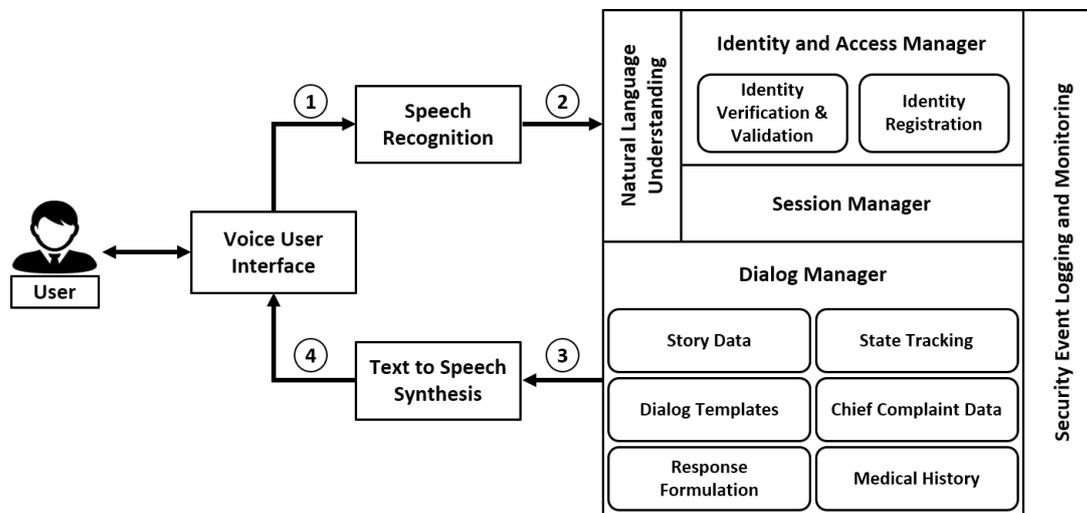


Figure 2. MIRA system architecture.

3.1. MIRA System Architecture

As illustrated in Figure 2, we added the identity and access manager, session manager, and security event logging and monitoring to the existing architecture of the virtual assistant [46,47], which overcomes the identified limitation of existing literature and virtual medical assistants. Here, the voice user interface provides an interactive communication medium between the MIRA and the user. We developed the prototype version of MIRA for Android due to wider compatibility with devices. Therefore, any smart devices (including smartwatches, smartphones, tablets, laptops, and some vendor-specific devices) that contain a microphone, speaker, and support Android can use the MIRA application. The speech recognition module recognizes human speech, then breaks it into voice samples, and transcribes each voice sample into text using the neural network algorithm for signal processing [48]. The MIRA speech recognition module automatically transcribes the voice sample in a context-specific format. Then the Natural Language Understanding (NLU) module determines the intent of user’s input based on the trained model. We used the Rasa framework for machine learning-based NLU and dialogue management [49]. For tokenization and part of speech annotation, we extracted the semantic concepts from the Unified Medical Language System (UMLS) [50]. The NLU also analyzes the nature of intent and forwards a request to a specific module (such as identity and access manager, session manager, or dialogue manager).

To the best of our knowledge, MIRA is the only virtual medical assistant that uses the concept of identity and access management [51]. We used our designed voice-based authentication protocol that identifies the user based on their voice samples [52]. Instead of random text, we matched the Mel-Frequency Cestrum Coefficients (MFCC) of each natural language input to provide a strong authentication mechanism. Moreover, the identity and access management consists of two sub-modules such as identity registration, and identity verification and validation. To use MIRA services, the user has to complete the registration process using the identity registration sub-module. For this purpose, MIRA collects a smart device identifier along with personal information such as name, address, gender, age, medical history, and voice samples. Among the collected information, the smart device identifier along with voice samples support authentication. The medical history, gender, and age help in the personalized recommendation. Moreover, this module also analyzes the collected information to avoid duplication and assigns a unique identifier of 7 digits, which can be used in a crisis such as authentication failure, identity verification, or permanent data removal. The identity

verification and validation sub-module verifies and validates the identity of a registered user. First, the smart device identifier links a user to the information that they provided during the registration phase. To authenticate the user, the smart device identifier helps to retrieve the provided voice sample MFCC; it is then compared with the calculated MFCC of natural language input to calculate the similarity index (SI). If the SI greater than 70%, then the user gets authenticated and MIRA generates an appropriate response.

The session manager assigns a session identifier to the authenticated users, which binds with the user identity and is valid for a specific session only. We used the keyword spotting technique, which detects 'Hello MIRA' and 'Bye MIRA' keywords in the spoken utterances. 'Hello MIRA' is used to initiate a session, and all the communication during this period is bound with the issued session identifier. The 'Bye MIRA' is used to terminate the ongoing session. We used two types of templates 'Hello [Given Name], How may I help you?' and 'Hello [Given Name], How may I help you today?' for greeting a new user with no medical history, and an established user with a medical history, respectively. Moreover, MIRA checks the validity of a corresponding session upon receiving an input request. In the case of timeout (idle for 60 minutes), the renewal request is forwarded to the session manager.

The dialogue manager is responsible for scenario understanding, state tracking and managing the flow of the conversation. This module identifies the conversational context from the natural language input and generates an appropriate response. It may be possible that the user starts another conversation without terminating or concluding the previous one. This type of conversation handling is not in the scope of this study. Moreover, the dialogue manager consists of six sub-modules. (i) The story data are used to train the dialogue management model. A story is the representation of a complete dialogue between the user and virtual assistant. We designed the story data manually from the recorded dialogue corpora that facilitate MIRA to make the conversation real and natural. (ii) The state tracking is the core module of MIRA that predicts the user goal (represented by slot-value pairs) at every dialogue turn. It maintains the conversation state, performs an action based on policy, and generates a relevant response after analyzing the natural language input. (iii) The dialogue templates consist of predefined statements that can be used by filling in the keyword. Although we trained a model to understand conversation and response generation, some statements are similar and common except for the keyword. Consider the statements 'Do you feel hungry?' and 'Do you feel tired?'. Both sentences are similar except for the keywords 'hungry' and 'tired'. To improve the performance and response generation of MIRA, we used templates for these kinds of statements that have similar semantics. (iv) The chief complaint data is the knowledge source that helps identify the conversation context. Based on the identified context, MIRA analyzes the dialogue corpora and asks a follow-up question. (v) The medical history consists of the health record that a user provided during registration. It also stores each recommendation along with the key attributes (sign and symptom) that results from the conversation between MIRA and the user. Keeping these health records helps the MIRA to generate a personalized decision for future conversations. (vi) The response formulation has a challenging role in the interaction because it generates a relevant response based on the input query. Therefore, this module takes the necessary information from different sub-modules of the dialogue manager and generates an appropriate text-based statement.

The text to speech synthesis analyzes and processes the text-based statement using natural language processing. Then, it converts the processed text into synthesized speech using digital signal processing and conveys it to the end-user in a polite female voice. MIRA deals with healthcare data and directs the user to a nearby appropriate medical specialist based on the chief complaints. This kind of dialogue contains sensitive information and its leakage may lead to serious consequences such as a masquerading and ransomware attacks. The security event logging and monitoring module continuously monitors the communication channels for anomalies. Also, it collects the information, which can be used as an audit trail for intrusion prevention and event management. With the proposed system architecture, MIRA understands single and multiple intent statements, supports adaptability, and provides data control.

3.2. Understanding the MIRA Digital Brain

According to [53], a virtual assistant consists of a digital brain, which is divided into a knowledge source, stock phrases, and conversation memory. Our state-of-the-art MIRA's digital brain is divided into a knowledge source, and stock phrases. We incorporated the conversation memory inside the stock phrases for efficient response generation. The knowledge source is an important part of a virtual assistant that helps in understanding the context of a conversation. Our proposed MIRA focused on the identification of a disease based on the user's chief complaint. In this regard, the first challenge that we faced was the selection of an appropriate dataset. We analyzed the publicly available datasets on the Internet, but to the best of our knowledge, none of the available datasets in English considered the patient chief complaint. Most of the datasets considered medical terminologies that are hard to understand for non-medical professionals. Therefore, we decided to create a dataset considering the patient chief complaints. For this purpose, we selected two well-known diseases, glaucoma and diabetes, due to the availability of collaborative medical specialists from the Yeouido Saint Mary's Hospital in the Republic of Korea. Under the hospital's legal policy (Institutional Review Board approval) and HIPAA (Health Insurance Portability and Accountability Act), we briefed the participants before their medical examinations, and a written consent form was signed by each participant. This form explained that the data would be collected anonymously and strictly used for research purposes (considered the privacy aspects) only. We collected 816 patient chief complaints and, based on the medical specialist's recommendation, classified them into glaucoma (48.5%), diabetes (46.2%), and other (5.3%). These labels were assigned based on the broad category of diseases. The glaucoma label consists of all patients, which includes angle-closure suspect, glaucoma suspect, and pure glaucoma patients. Similarly, the diabetes label consists of all types of diabetic patients, which include type 1, type 2, and gestational. The other label consists of those patients that have diseases except glaucoma and diabetes, including normal conditions. We represented the data in tabular form that consist of 816 rows and 32 columns. Each row represents one patient with potential symptoms, while the columns represent observed features for that patient, including the class of diagnosis label (glaucoma, diabetes, or other). Table 2 describes 31 features of the MIRA dataset. Collecting such data helps us to identify specific patients based on their chief complaints since the categorization of these patients is based on different laboratory test results and medical specialist opinion.

After the creation of the knowledge source, the next challenge was to identify the most appropriate predictive model. For this purpose, we used MOD [54], which filters out seven applicable machine learning models (including decision trees, naive Bayes, K-nearest neighbors, random forest, random tree, decision stump, and deep learning) based on the provided dataset features. To determine the accuracy of each predictive model for MIRA's dataset, we used RapidMiner with 10-fold cross-validation and evaluated the predictive model accuracy as shown in Figure 3. The result shows higher accuracy for the deep learning model (99.14%) because it learns from data incrementally and identifies the hidden relationships. Therefore, we selected deep learning as the best suitable predictive model for MIRA. The predictive model along with knowledge source helps in context identification of a dialogue corpus, which determines the category of the disease such as glaucoma, diabetes, or other. The stock phrases help MIRA to understand the user intent (what the user is trying to say) and support response generation. We searched online for publicly available patient-doctor dialogue corpora in the English language, but none of the relevant datasets were found. Therefore, we decided to design the dialogue corpus from the recorded patient-doctor conversation, which includes 816 dialogue corpora (11,532 utterances). We manually annotated each utterance for NLU and the dialogue manager to make the interactive environment of MIRA as real and natural as possible.

Table 2. MIRA Dataset features with ranges, measurement units, and meaning of each feature.

Feature Name	Value Range	Measurement Unit	Meaning
Age	[17, 73]	Years	Age of the patients
Gender	0, 1	Category	Male or Female
Urinating often	0, 1	Boolean	Frequent urination can be a symptom of many diseases such as diabetes
Feeling thirsty	0, 1	Boolean	Urge to drink too much may indicate diseases such as diabetes
Feeling hungry	0, 1	Boolean	Patient may feel strong hunger due to low blood sugar; it may indicate diabetes because of an abnormal glucose level
Extreme fatigue	0, 1	Boolean	Uncontrolled blood glucose may leads to tiredness
Blurry vision	0, 1	Boolean	In diabetes, a high blood glucose level may lead to temporary blurring of eyesight; moreover, damaged optic nerves increase the intraocular pressure that may leads to haziness or blurry vision
Slow-healing wounds	0, 1	Boolean	High blood glucose level may affect the blood circulation, which may leads to the slow-healing of wounds.
Weight loss	0, 1	Boolean	The body starts burning fat and muscle for energy with insufficient insulin
Has tingling sensation	0, 1	Boolean	Diabetic neuropathy may lead to tingling sensations in fingers, toes, hands, and feet, and burning may occur as well
Pain	0, 1	Boolean	Diabetic neuropathy may leads to pain in different body parts such as arms, legs or sometimes the whole body
Numbness of hands	0, 1	Boolean	Diabetic neuropathy may lead to numbness of hands
Numbness of foot	0, 1	Boolean	Diabetic neuropathy may lead to numbness of feet
Burning sensation in eye	0, 1	Boolean	Stinging or irritating sensation in the eyes
Color vision impairment	0, 1	Boolean	Color vision impairment is the initial symptom of glaucoma
Difficulty walking	0, 1	Boolean	Glaucoma patients frequently complain of difficulty walking
Difficulty in stair climbing	0, 1	Boolean	Glaucoma patients frequently complain of difficulty climbing stairs
Difficulty in face recognition	0, 1	Boolean	Glaucoma patients frequently complain of difficulty recognizing faces
Difficulty driving	0, 1	Boolean	Glaucoma patients frequently complain of difficulty driving
Double vision	0, 1	Boolean	Diplopia is considered to be a warning for glaucoma
Dryness of eyes	0, 1	Boolean	Dryness of eyes is due to the lack of proper tear production
Swelling of eyelids	0, 1	Boolean	Occurs due to inflammation or excess of fluid
Tear in eyes with a strong glare	0, 1	Boolean	Unusual squinting or blinking due to a strong glare or light
Image quality decrease	0, 1	Boolean	Peripheral vision loss may be an early symptom of glaucoma
Itchiness	0, 1	Boolean	Itchiness caused due to the low quantity of eye fluid or low interocular pressure
Nausea and vomiting	0, 1	Boolean	Severe eye pain may cause nausea and vomiting
Headache	0, 1	Boolean	Severe eye pain may cause headache
Night blindness	0, 1	Boolean	Nyctalopia is a condition where the eye is unable to adapt to the surrounding conditions such as low-light, or nighttime
Redness of eyes	0, 1	Boolean	Caused due to swollen or dilated blood vessels
Severe eye pain	0, 1	Boolean	The rapid eye pressure increase causes severe eye pain
Sudden onset of visual disturbances usually in low light	0, 1	Boolean	The basic signs and symptoms of acute angle closure glaucoma

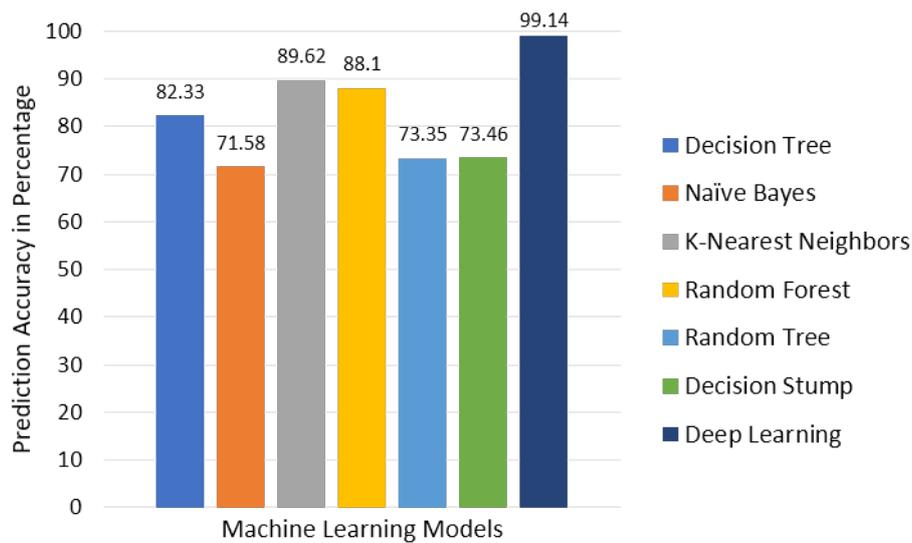


Figure 3. MIRA prediction accuracy with machine learning models.

3.3. Case Study

To understand the working scenario of MIRA, consider John Doe, a registered user, who wants to discuss his medical condition with MIRA and is looking for an appropriate medical specialist nearby. John started the conversation by saying 'Hello MIRA'. The speech recognition recognizes the natural language input as received from the voice user interface, transcribes it into text, and sends it to the NLU to identify the intent of the utterance, which is a greeting in this case. The intent-text pair (intent: greeting, Text: Hello MIRA) along with voice-print is sent to the identity and access manager, which verifies and validates John's identity using the MFCC matching technique. Upon approval, the request is forwarded to the session manager, which determines whether John has an ongoing session or the phrase is to initiate a new conversational session. According to the session manager, John does not have an ongoing session. Therefore, the session manager generates a new session identifier linked with John's identity and forwards the request to the dialogue manager for generating a relevant response. At first, the dialogue manager analyzes the state for an ongoing conversation using state tracking, then infers the intent of the request based on the chief complaint data and medical history. In this case, John did not provide any medical history during the registration phase and initiated the conversation with a greeting utterance, which does not link to any of the chief complaints. Therefore, the inferred request is forwarded to the story data for selecting an appropriate story. Then, a new user greeting template is selected using the dialogue templates and is forwarded to the response formulation, which customizes the template based on the user identifier to generate a text statement. The text to speech synthesis receives this text, transcribes it into a spoken response, and plays it on a smart device speaker. A similar procedure will be followed for handling each dialogue corpus. At any point of the conversation, the user can say 'Bye MIRA' to terminate the session. Figure 4 illustrates the MIRA implementation model for handling a complex conversation. The different colored lines present the workflow and inter-connectivity between the basic modules.

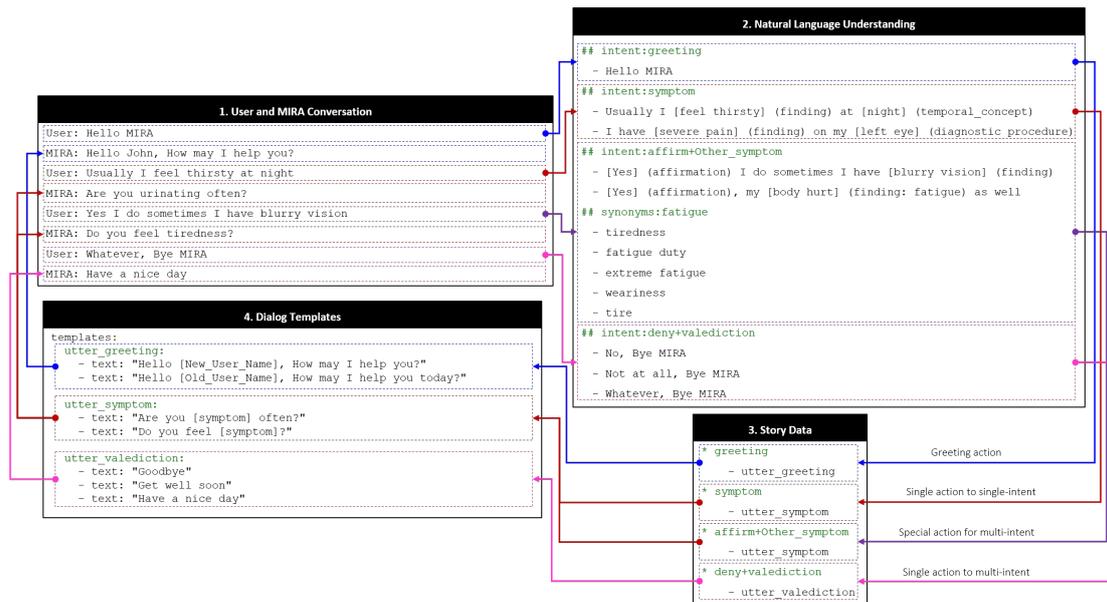


Figure 4. MIRA implementation model for conversational handling.

Figure 5 presents the MIRA smartphone application screenshots. The user interface shows a circularly shaped gray button on the main screen, which can be used to activate MIRA by pressing the button. Upon activation, MIRA starts listening, and the color of the button changes to bright green. We set the listening duration to 5 seconds, but it can be changed to 1 minute from the application setting. When the time is up, MIRA starts analyzing the spoken natural language and changes the button color to orange. We used the color change technique because warm colors have a positive impact on the user’s emotions and behavior as per psychology [55]. Furthermore, MIRA displays the input and output natural language on the smartphone screen in the form of a chat bubble for better understanding along with the spoken response. MIRA switches to an idle state (gray color), if the user does not speak for 5 seconds, which requires reactivation by pressing the gray button. However, the session identifier will be valid until the user terminates by saying ‘Bye MIRA’ or the conversation is idle for 60 consecutive minutes. As a final recommendation, a frame of Google maps shows the nearby appropriate medical specialist. By clicking the map frame, the query will open in Google maps.

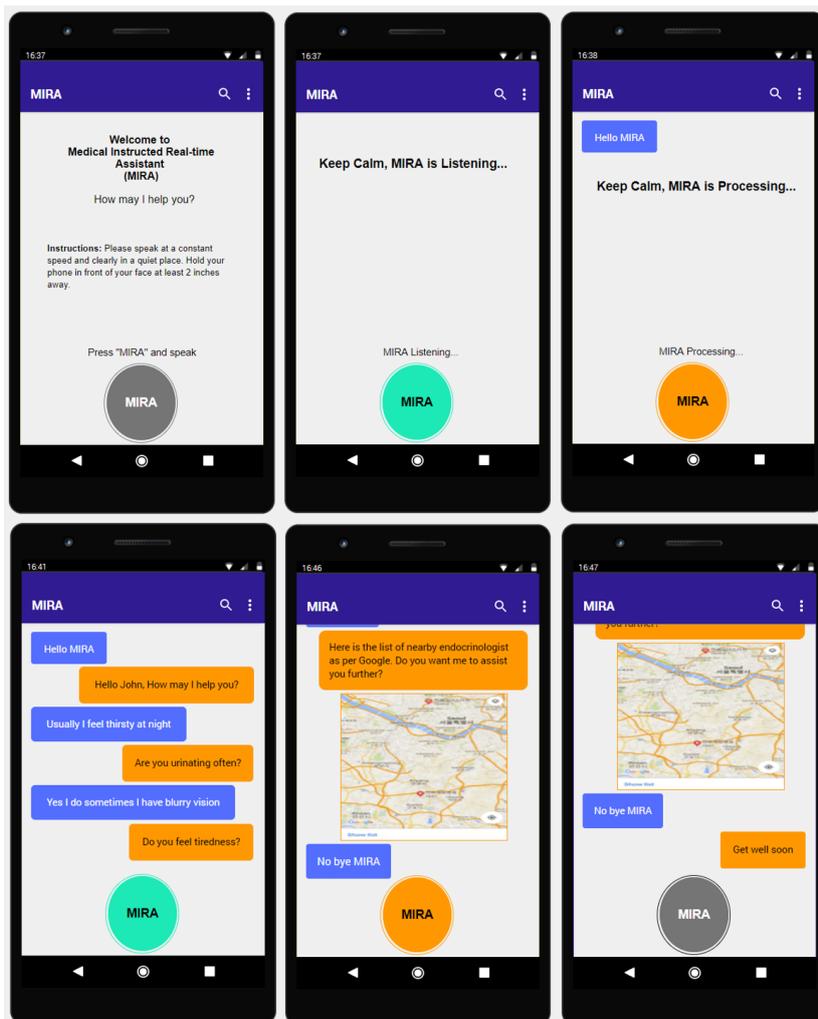


Figure 5. MIRA smartphone application screenshot on android pie.

4. Evaluation

MIRA provides efficient and reliable healthcare services to the users. To ensure productivity, we evaluated MIRA based on performance measures, task completion, security, and user experience. For this purpose, we circulated a call for participants on the university's mailing list and social media. A total of 33 participants belonging to seven countries registered, including 20 males and 13 females within the age group of 18 to 43 years as shown in Figure 6. The participants were affiliated with different departments such as Healthcare Subject Matter Experts (5), Medical Practitioners (4), and students belonging to Medical (7), Computer Science (9), Bioinformatics (3), Life Science (3), and International Relations (2) disciplines. Each participant was given a set of procedural documents, which contained a checklist of tasks, consent form, hints for acting as a particular patient type, and a user experience questionnaire. The consent form clearly describes the data collection procedure, including audio and video recording of interactions with MIRA, data storage, data usage, and disposal details. Moreover, participants were shortly briefed about the goal of the activity, and we instruct them to sign the consent form after reading it carefully. Upon agreement, the voice sample along with demographic information (name, address, gender, age, and medical history) was collected to complete the MIRA registration process.

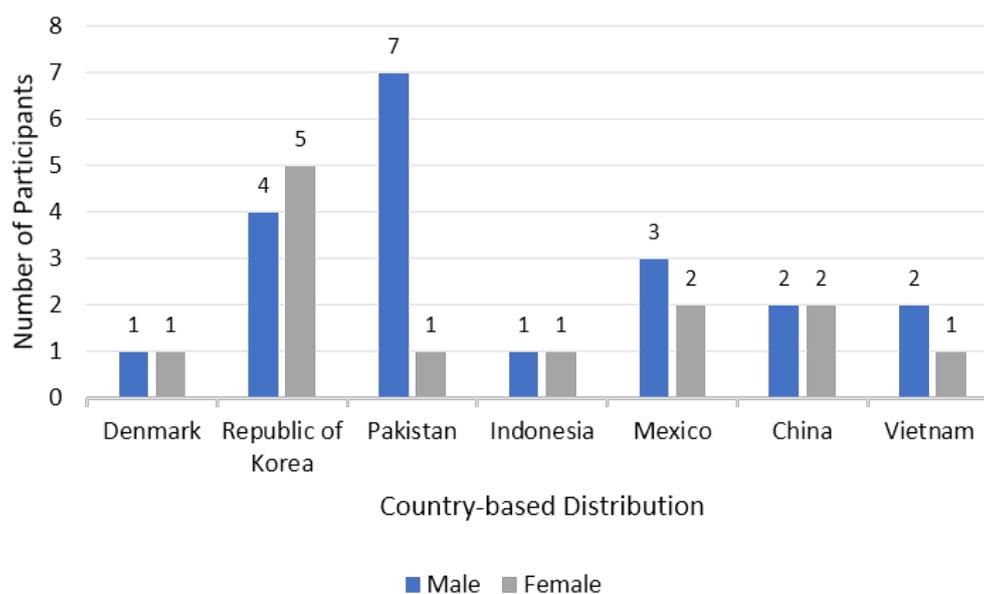


Figure 6. Country-based distribution of MIRA evaluation participants.

As per the scope of this study, MIRA predicts glaucoma and diabetes based on the trained model. The remaining diseases including normal conditions are out of scope and are considered under the other label. Therefore, MIRA analyzed user interactions, identified chief complaints, and categorized these as glaucoma, diabetes, or other. Among the 33 registered participants, 17 did not belong to the medical profession. For this reason, we provided a list of chief complaints as described in Table 3, which guide the participants to act as a patient for three health conditions. In the case of other label, we selected cardiovascular and orthopedic chief complaints that are similar to glaucoma and diabetes. If MIRA did not generate a final recommendation for some reason, then it politely responded 'I am sorry, I am not able to diagnose your disease based on the provided knowledge. Do you want me to assist you further?'. Moreover, the participants were allowed to use synonyms, ask questions in a random sequence, and interact in a natural way of communication.

4.1. Experimental Setup

We set up an interactive environment based on the availability of resources, which includes three android smartphones (Samsung Galaxy S7), three iPhones (6s), three cell phone holders, and three tripod mounts. The MIRA application was installed on three android smartphones, and these were attached to classroom desks with the help of cell phone holders, which can be adjusted. The three iPhones attached to tripod mounts were used for audio and video recording of each user's interaction with MIRA. Complete sets of equipment were placed at three corners of the classroom, which include an android smartphone, cell phone holder, iPhone, tripod mount, classroom desk, and chair. Only three participants can interact with MIRA simultaneously in the design experimental setup. Therefore, we divided the participants into 11 groups (three members per group) based on their availability and feasibility. Each member of the group can interact with MIRA independently for an allocated time of 60 minutes while acting like a patient using the provided hints.

4.2. Performance Evaluation

To assess the effectiveness of MIRA, we used the common performance evaluation measure based on an independently distributed confusion matrix, as described in Table 4. The values were assigned based on the final recommendation label. The diagonal and off-diagonal values of the confusion matrix present correctly classified and incorrectly classified results, respectively. Similarly, the rows

and columns of the confusion matrix show actual values per label and predicted value per label, respectively. Furthermore, the characteristics of performance evaluation measurement are reflected in terms of accuracy, precision, sensitivity, specificity, and F-measure. The corresponding description along with formulas for these measures are described as follows. Each participant completed the interaction for three health conditions that included glaucoma, diabetes, and other. Figure 7 illustrates the value of each label. We recorded a total of 99 dialogue corpus based on the interactions of 33 participants.

Table 3. Sample of hints for acting as glaucoma, diabetes, and other, patient types.

Chief Complaints of Different Diseases (Feel Free to Use Any Synonyms Related to These Chief Complaints)		
Glaucoma	Diabetes	Other
Blurry vision	Blurry vision	Sweating
Burning sensation or dryness or itchiness in eye(s)	Extreme fatigue	Pain
Color vision impairment	Feeling very hungry	Nausea and vomiting
Difficulty in driving, face recognition, stair climbing, and walking	Feeling very thirsty	Shortness of breath
Double vision or decrease in image quality or sudden onset of visual disturbance usually in low light	Numbness of feet	Discomfort in body parts such as neck, jaw, shoulder, upper back, or abdominal
Nausea and vomiting with headache	Numbness of hands	Unusual fatigue
Night blindness	Pain	Lightheadedness or dizziness
Red eyes	Slow healing of cuts and bruises	Stiffness
Severe eye pain	Tingling sensation	Swelling
Swelling in eyelid	Urinating often	Instability
Tears in eyes with a strong glare	Weight loss	Deformity

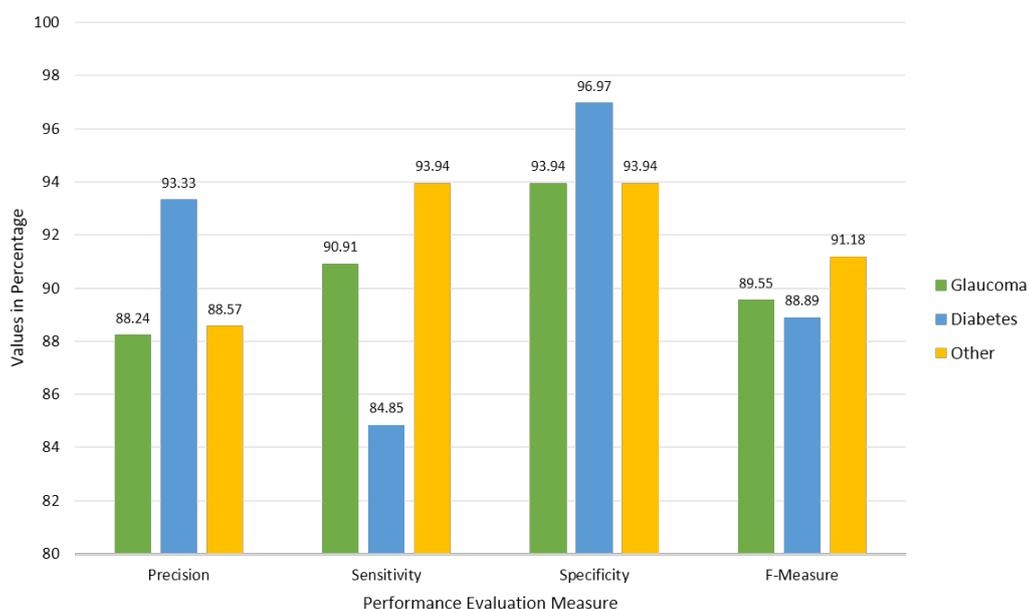


Figure 7. Performance evaluation measure of interactive scenarios.

Table 4. MIRA confusion matrix.

	Glaucoma	Diabetes	Other
Glaucoma	30	1	2
Diabetes	3	28	2
Other	1	1	31

- Accuracy identifies the effectiveness of an algorithm based on the probability of true values as stated in Equation (1). MIRA gets an overall accuracy of 89.8% because it correctly identified 90.9% glaucoma (30), 84.8% diabetes (28), and 93.9% other (31) labels among the recorded dialogue corpus (99).

$$Accuracy = \frac{SumofCorrectlyClassified}{TotalNumberofClassification} \quad (1)$$

- Precision or confidence presents the positive predictive value of a label that can be derived using Equation (2). We obtained the precision for each label including glaucoma (88.24%), diabetes (93.33%), and other (88.57%), with an average precision of 90%.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (2)$$

- Sensitivity (also known as recall) corresponds to the true positive rate of a specific label and can be computed with Equation (3) for glaucoma (90.91%), diabetes (84.85%), and other (93.94%), with an average value of 89.8%.

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative} \quad (3)$$

- Specificity corresponds to the true negative rate and can be computed using Equation (4) for a specific label for glaucoma (93.94%), diabetes (96.97%), and other (93.94%), with an average value of 94.9%.

$$Specificity = \frac{TrueNegative}{FalsePositive + TrueNegative} \quad (4)$$

- The F-measure, also known as F-score or F1-score, is the weighted harmonic mean of precision and sensitivity (recall) as stated in Equation (5). The F-Measures for each label in MIRA were as follows: glaucoma (89.55%), diabetes (88.89%), and other (91.18%), with an average value of 89.8%. We used $\beta = 1$ that evenly balances the F-score based on precision and sensitivity.

$$F - Measure = \frac{(\beta^2 + 1) * Precision * Sensitivity}{\beta^2 * Precision + Sensitivity} \quad (5)$$

4.3. Task Completion

Task completion is an important factor in the virtual assistant. It measures the task success probability of dialogue corpora. To assess MIRA's task completion, we used the PARADISE (PARAdigm for DIalogue System Evaluation) framework that uses the Kappa coefficient to operationalize the measure of task-based success [56]. The Kappa coefficient k measures the success rate of task completion and is computed with Equation (6).

$$k = \frac{P(A) - P(E)}{1 - P(E)} \quad (6)$$

$P(A)$ is the proportion of times that agreement occurs between the actual and scenario attribute value. $P(E)$ is the proportion of times when the agreement between the actual and scenario attribute value is expected. The value of k considers task complexity and assesses the virtual assistant by correcting for the expected agreement and performing different tasks. If agreement is only expected by chance, then $k = 1$ and $k = 0$ for total agreement and no agreement, respectively. Moreover, if the expected chance of agreement ($P(E)$) is unknown, then it can be calculated from the confusion matrix using Equation (7).

$$P(E) = \sum_{i=1}^n \left(\frac{t_i}{T} \right)^2 \quad (7)$$

Here, t_i is the sum of the i^{th} column frequency of the confusion matrix. T is the sum of frequency $t_1 + t_2 + \dots + t_n$ in the confusion matrix. Similarly, $P(A)$ can be calculated from the confusion matrix with Equation (8), if unknown.

$$P(A) = \frac{\sum_{i=1}^n M(i, i)}{T} \quad (8)$$

MIRA task completion based on the PARADISE framework gives Expected Agreement $P(E) = 0.334$, Actual Agreement $P(A) = 0.898$, and Kappa Coefficient $k = 0.848$. The interpretation of Kappa categorized MIRA as ‘Almost Perfect’ in term of task completion [57].

4.4. Security

Healthcare applications deal with sensitive data such as medical records, health conditions, and quality of life. The illegal usage of these data may lead to several attacks. For this purpose, we launched a masquerading attack based on the scope of this study. The prevention of masquerading attack also minimizes the risk of ransomware.

The masquerading attack uses a fake identity to gain unauthorized access [58]. To launch this attack on MIRA, we asked the members of each group to shift their positions. Suppose the participants on positions C, B, and A will shift to A, C, and B respectively. Then, the adjacent member gets access to the authenticated user account of MIRA and starts an interaction. During the analysis of natural language input, MIRA verifies the device identifier, but is unable to validate the MFCC value. Therefore, MIRA holds the ongoing session and asks the participant for identity verification as ‘Sorry for the interruption, malicious activity was detected. To proceed with the ongoing session, please enter your seven digit identity verification key’. At this stage, the user has to enter the identity verification key to interact with MIRA. Moreover, if an unauthorized user wants to interact after session time out (60 minutes), then MIRA will respond as ‘I am sorry, but I am not able to verify your identity. Do you want me to assist you through the registration process?’. Furthermore, one smart device identifier can bind with multiple user identities, which means that more than one user can use the same device, but registration is mandatory for each user. The results show that MIRA prevents against masquerading attacks because none of the participants were able to interact with other user applications due to voice-based authentication.

4.5. User Experience

After interacting with MIRA, the participants were asked to fill out the User Experience Questionnaire (UEQ) [59], which covers all the aspect of user experience in a comprehensive way. The UEQ is widely used as a subjective measurement of user experience and provides a data analysis tool for assessing user responses. Therefore, we used it to evaluate the MIRA user experience. It consists of 26 items using a 7 point Likert scale for rating. The results of these 26 items are mapped with 6-dimensional scales such as attractiveness (6 items), perspicuity (4 items), efficiency (4 items), dependability (4 items), stimulation (4 items), and novelty (4 items) as shown in Figure 8. The x -axis and y -axis present the list of items and rating scales (extremely good(+3), neutral (0), horribly bad (−3)),

respectively. Furthermore, the 6-dimensional scales are grouped into pragmatic quality (perspicuity, efficiency, and dependability), and hedonic quality (stimulation, originality). The pragmatic deals with task-related quality aspects, while the hedonic describes non-task related quality aspects.

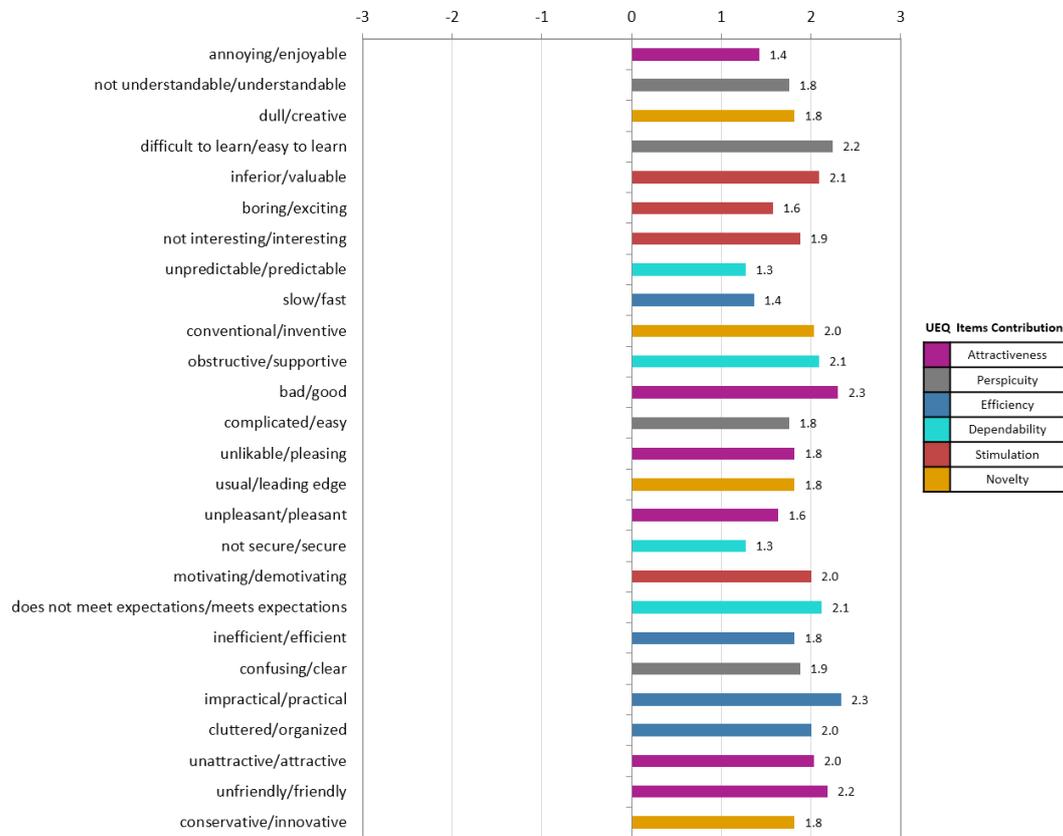


Figure 8. MIRA user experience questionnaire mean value per item.

Figure 9 illustrates the result of MIRA based on 6-dimensional scales, which exhibit accurate measurements because the values are greater than 1.6. Moreover, Figure 10 presents MIRA’s attractiveness and pragmatic quality along with hedonic quality, where the value is greater than 1.80, reflecting a positive evaluation based on UEQ criteria. To identify the correlation of items per scale, UEQ uses Cronbach’s alpha-coefficient, which measures the consistency of a scale as shown in Table 5. The value of attractiveness is higher than 0.7, which means that all users enjoyed the interactions with MIRA. Most of the participants recommend an avatar instead of a simple user interface for MIRA. Therefore, the alpha-coefficient value of novelty was less than 0.5. Furthermore, Figure 11 presents a comparative analysis of MIRA based on the UEQ benchmark dataset, which consists of 401 product evaluations collected from 18483 participants. The results show that MIRA is relatively good in all aspects based on the benchmark data.

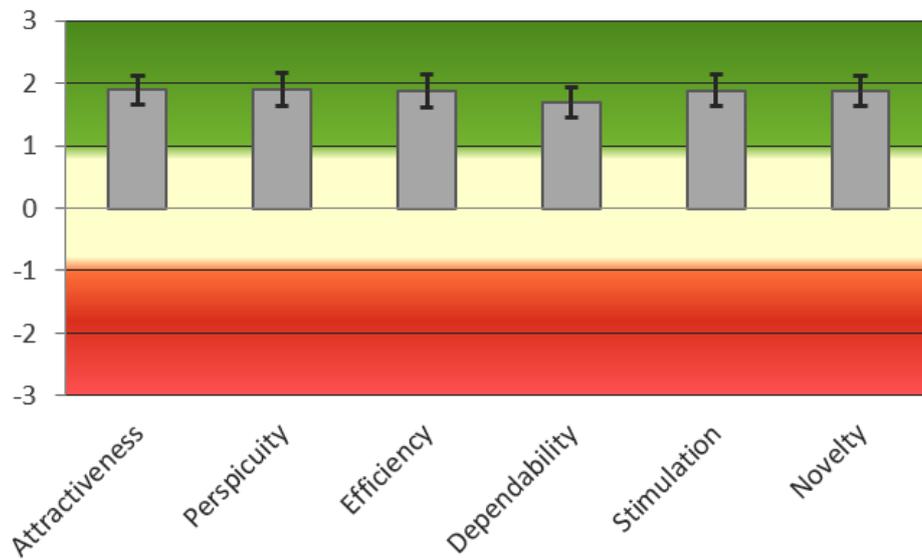


Figure 9. MIRA user experience questionnaire resulting scores on six dimensional scale.

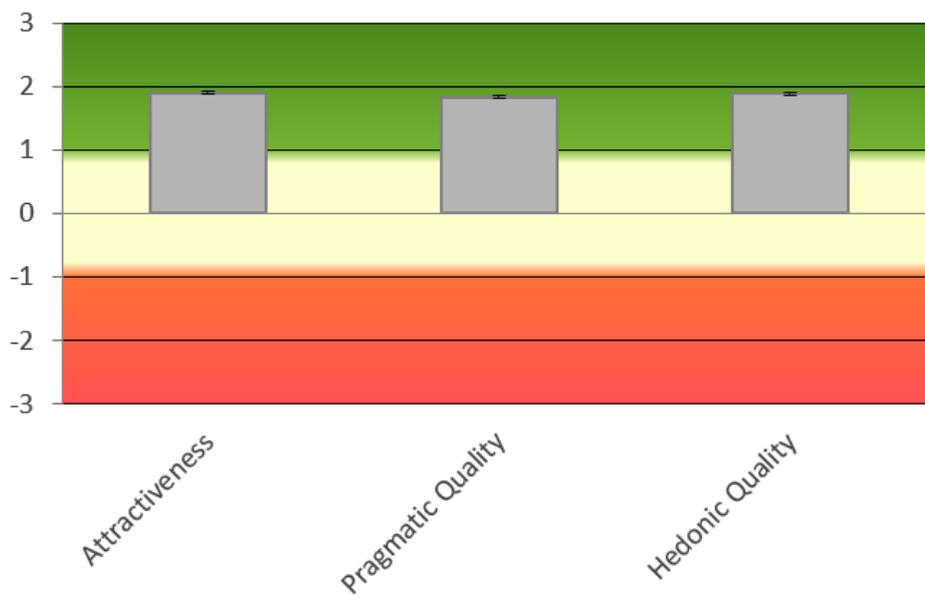


Figure 10. MIRA user experience questionnaire aggregated score of pragmatic and hedonic qualities.

Table 5. Correlation of items per scale using Cronbach’s Alpha Coefficient.

Scale	Alpha-Coefficient
Attractiveness	0.74
Perspicuity	0.67
Efficiency	0.77
Dependability	0.60
Stimulation	0.67
Novelty	0.48

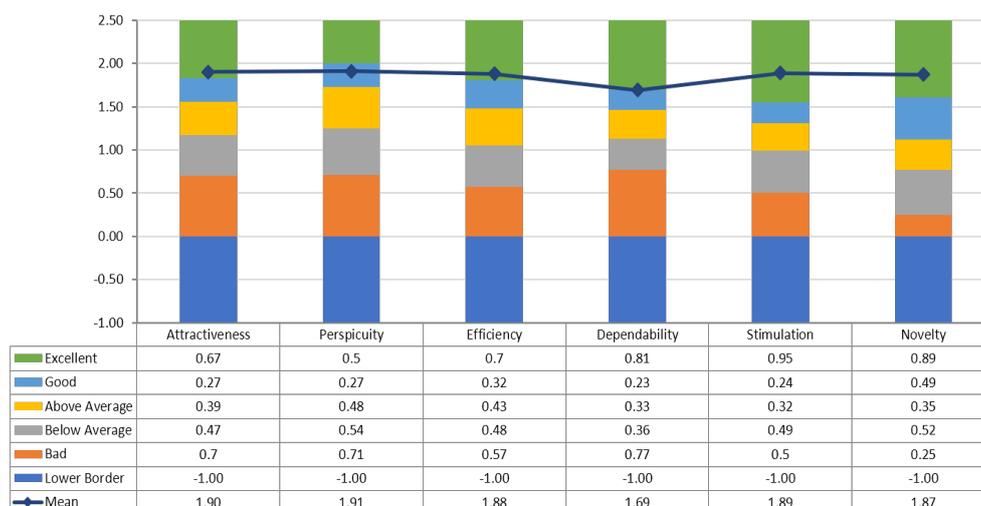


Figure 11. MIRA user experience questionnaire scores on six dimensions scales along with benchmark data.

4.6. Discussion

MIRA was evaluated by 33 participants belonging to different domains, age groups, genders, and diverse nationalities. The participants were given 60 minutes to complete a list of tasks during the interaction with MIRA. Among the 33 participants, 27 completed their tasks at an average time of 40 minutes because their interaction was smooth with little or no misinterpretation. However, 6 participants took an average of 55 minutes due to several misinterpretations such as ‘thirsty’ as ‘thirty’, ‘tired’ as ‘tire’, ‘driving’ as ‘diving’, and ‘tear’ as ‘tire’. Based on these interactions, MIRA gets an overall accuracy of 89% because it used the deep learning predictive model, which learns from the data incrementally and manages complex dialogues efficiently. We considered the macro-average instead of micro-average for calculating precision (90%), sensitivity (89.8%), specificity (94.9%), and F-measure (89.8%) of the complete system. Please note that the macro-average gives equal weight to each class label, while the micro-average results are biased towards the larger class label. Therefore, we showed impressive results for MIRA in terms of efficiency and effectiveness. Moreover, the PARADISE framework was used to evaluate the task completion of MIRA, where the actual agreement ($P(A) = 0.898$) is better than agreement-by-chance ($P(E) = 0.334$). Because the stock phrases were designed from real conversations that facilitated MIRA for a better understanding of natural language input. The Cohen’s Kappa value ($k = 0.848$) was interpreted as ‘Almost Perfect’ because MIRA generated the response in a real and natural way using a female voice that keeps the user motivated to continue the interaction.

MIRA also keeps a record of the conversational dialogue corpus along with final recommendation about the appropriate medical specialist that supports personalized interactions with the established user. For the prototype version of MIRA, we considered authentication instead of confidentiality, integrity, and availability. A strong authentication mechanism minimizes the risk of exploiting security vulnerabilities but will affect the performance and efficiency of the system. Therefore, we used the lightweight version of our designed voice-based authentication protocol, which identifies the user based on the extracted MFCC value of natural language utterances; this method was evaluated for masquerading attack. The results showed that MIRA successfully identified the user in real time based on their voice samples and strongly resisted a masquerading attack.

We used UEQ for evaluating the user experience because it provides ease of data analysis and calculates the necessary statistics accordingly. Due to reliability, different organizations used UEQ for evaluating their products and consider it to be a good measure. According to UEQ, MIRA was evaluated in terms of attractiveness, pragmatic quality, and hedonic quality, where the value of pragmatic is smaller than the other two qualities (attractiveness and hedonic). This is due to the low value of secure and predictable items under the category of pragmatic quality because some

participants considered secure in terms of security, but it evaluates the user's feelings regarding the interaction control. Moreover, MIRA uses synonyms of specific words for generating a relevant response, which may be unpredictable for a conversational scenario in some situations. Suppose in one interaction MIRA asked a user, 'How about your empty-bellied?' instead of 'Do you feel extremely hungry?'. The value of pragmatic quality is affected by these two factors. However, the overall results of UEQ present positive feedback, and users were satisfied with MIRA's interactive communication.

After the completion of tasks, the participants were awarded a shopping coupon worth 30,000 KRW as an incentive. The participants belonged to diverse nationalities that helped assess how MIRA deals with a variety of accents as well. According to our analysis, some participants do not realize the voice-based authentication mechanism due to the lightweight protocol until they were asked to switch their positions for performing a masquerading attack. In the future we plan to evaluate MIRA with real glaucoma and diabetic patients, then compare the results of both assessments. Furthermore, we will evaluate MIRA for relevant emerging cyber-attacks.

5. Conclusions

In this study, we introduced a state-of-the-art virtual medical assistant, MIRA, that interacts with the user in a spoken natural language, diagnoses a disease based on a user's chief complaint, and refers the user to a nearby appropriate medical specialist. The key contribution of MIRA includes disease identification based on chief complaint, understanding single and multiple intents, a voice-based authentication mechanism, conversational state tracking, and continuous monitoring of the system for detecting anomalies. Moreover, we designed a chief complaint dataset and stock phrases from the recorded dialogue corpora. MIRA is the first assistant of its kind that considers security aspects (such as authentication), which requires improvements in terms of transmission security and audit control to become HIPAA compliant. The designed knowledge source of MIRA considered glaucoma and diabetes chief complaints only, which can be extended to other medical conditions in the future.

There are many challenges in developing these kinds of interactive systems such as privacy concerns, accuracy constraints, correct decision making, precise response generation, and gaining user trust. The compliance with standards may help in risk minimization. Besides these challenges, it is beneficial for society, especially in underdeveloped countries, where people are suffering from many diseases due to the lack of healthcare facilities. These kinds of virtual medical assistants help the patient identify an appropriate medical specialist and reduce healthcare cost. Also, it supports medical practitioners and students in clinical decision making.

Author Contributions: U.U.R. is the principal researcher, who proposed the idea, designed and developed the prototype version, conducted the experiments based on the designed scenarios, and wrote the paper. D.J.C. and Y.J. provided the medical related information, supported in data acquisition and analysis. U.A. and M.A.R. contributed to participant management, English proofreading and finalized content flow in the manuscript. S.L. supervised the whole process, provided advisory feedback, and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-0-01629) supervised by the IITP (Institute for Information & communications Technology Promotion). This work was supported by the Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00655). This work was also supported by the National Research Foundation (NRF) under the NRF-2016K1A3A7A03951968 and NRF-2019R1A2C2090504.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Canbek, N.G.; Mutlu, M.E. On the track of artificial intelligence: Learning with intelligent personal assistants. *J. Hum. Sci.* **2016**, *13*, 592–601. [[CrossRef](#)]
2. Van Os, M.; Saddler, H.J.; Napolitano, L.T.; Russell, J.H.; Lister, P.M.; Dasari, R. Intelligent Automated Assistant for TV User Interactions. US Patent 9,338,493, 2016.

3. Bartie, P.; Mackaness, W.; Lemon, O.; Dalmas, T.; Janarthnam, S.; Hill, R.L.; Dickinson, A.; Liu, X. A dialogue based mobile virtual assistant for tourists: The SpaceBook Project. *Comput. Environ. Urban Syst.* **2018**, *67*, 110–123. [CrossRef]
4. Page, L.C.; Gehlbach, H. How an artificially intelligent virtual assistant helps students navigate the road to college. *AERA Open* **2017**, *3*, 2332858417749220. [CrossRef]
5. Lam, M.S. Keeping the Internet Open with an Open-Source Virtual Assistant. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*; ACM: New York, NY, USA, 2018; pp. 145–146.
6. Austerjost, J.; Porr, M.; Riedel, N.; Geier, D.; Becker, T.; Scheper, T.; Marquard, D.; Lindner, P.; Beutel, S. Introducing a Virtual Assistant to the Lab: A Voice User Interface for the Intuitive Control of Laboratory Instruments. *SLAS TECHNOL. Transl. Life Sci. Innov.* **2018**, *23*, 476–482. [CrossRef]
7. Yan, R.; Song, Y.; Wu, H. Learning to respond with deep neural networks for retrieval-based human-computer conversation system. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*; ACM: New York, NY, USA, 2016; pp. 55–64.
8. Hwang, E.J.; Jung, J.Y.; Lee, S.K.; Lee, S.E.; Jee, W.H. Machine Learning for Diagnosis of Hematologic Diseases in Magnetic Resonance Imaging of Lumbar Spines. *Sci. Rep.* **2019**, *9*, 6046. [CrossRef] [PubMed]
9. Omondigbe, D.A.; Veeramani, S.; Sidhu, A.S. Machine Learning Classification Techniques for Breast Cancer Diagnosis. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2019; Volume 495, p. 012033.
10. Pignoni, A.; Delvecchio, G.; Madonna, D.; Bressi, C.; Soares, J.; Brambilla, P. Can Machine Learning help us in dealing with treatment resistant depression? A review. *J. Affect. Disord.* **2019**, *259*, 21–26. [CrossRef]
11. Künzel, S.R.; Sekhon, J.S.; Bickel, P.J.; Yu, B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4156–4165. [CrossRef]
12. Callahan, A.; Shah, N.H. Machine learning in healthcare. In *Key Advances in Clinical Informatics*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 279–291.
13. Sinsky, C.; Colligan, L.; Li, L.; Prgomet, M.; Reynolds, S.; Goeders, L.; Westbrook, J.; Tutty, M.; Blike, G. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann. Intern. Med.* **2016**, *165*, 753–760. [CrossRef]
14. Nuance AI-Powered Virtual Assistants for Healthcare. Available online: <https://www.nuance.com/healthcare/ambient-clinical-intelligence/virtual-assistants.html> (accessed on 13 March 2019).
15. Suki Let Doctors Focus on What Matters. Available online: <https://www.suki.ai/about-us> (accessed on 23 March 2019).
16. Robin Healthcare. Available online: <https://www.robinhealthcare.com> (accessed on 24 March 2019).
17. UHS Drives Quality through Cloud Speech and CDI Workflow. Available online: https://www.nuance.com/content/dam/nuance/en_us/collateral/healthcare/case-study/cs-uhs-en-us.pdf (accessed on 15 March 2019).
18. Plastic Surgery Specialist Reduces Time Per Patient Note. Available online: <https://resources.suki.ai/home/case-study-dr-ereso-plastic-surgeon> (accessed on 24 March 2019).
19. Plastic Surgery Specialist Reduces Time Per Patient Note. Available online: <https://www.mobihealthnews.com/news/north-america/voice-enabled-clinician-workflow-tool-robin-healthcare-raises-115m> (accessed on 2 October 2019).
20. Medwhat Virtual Medical Assistant. Available online: <https://medwhat.com/> (accessed on 2 April 2019).
21. Your.MD Symptom Checker. Available online: <https://www.your.md/> (accessed on 2 April 2019).
22. Sensely Engage Your Members. Reduce Your Costs. Available online: <https://www.sensely.com/> (accessed on 2 April 2019).
23. Bickmore, T.W.; Trinh, H.; Olafsson, S.; O’Leary, T.K.; Asadi, R.; Rickles, N.M.; Cruz, R. Patient and consumer safety risks when using conversational assistants for medical information: An observational study of Siri, Alexa, and Google Assistant. *J. Med. Internet Res.* **2018**, *20*, e11510. [CrossRef]
24. Semigran, H.L.; Linder, J.A.; Gidengil, C.; Mehrotra, A. Evaluation of symptom checkers for self diagnosis and triage: Audit study. *BMJ* **2015**, *351*, h3480. [CrossRef]
25. Crestani, F.; Du, H. Written versus spoken queries: A qualitative and quantitative comparative analysis. *J. Am. Soc. Inf. Sci. Technol.* **2006**, *57*, 881–890. [CrossRef]
26. Philip, P.; Bioulac, S.; Sauteraud, A.; Chaufton, C.; Olive, J. Could a virtual human be used to explore excessive daytime sleepiness in patients? *Presence Teleop. Vir. Environ.* **2014**, *23*, 369–376. [CrossRef]

27. Philip, P.; Micoulaud-Franchi, J.A.; Sagaspe, P.; De Sevin, E.; Olive, J.; Bioulac, S.; Sauteraud, A. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Sci. Rep.* **2017**, *7*, 42656. [[CrossRef](#)] [[PubMed](#)]
28. Tanaka, H.; Negoro, H.; Iwasaka, H.; Nakamura, S. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLoS ONE* **2017**, *12*, e0182151. [[CrossRef](#)] [[PubMed](#)]
29. Dimeff, L.A.; Jobs, D.A.; Chalker, S.A.; Piehl, B.M.; Duvivier, L.L.; Lok, B.C.; Zalake, M.S.; Chung, J.; Koerner, K. A novel engagement of suicidality in the emergency department: Virtual Collaborative Assessment and Management of Suicidality. In *General Hospital Psychiatry*; Elsevier: Amsterdam, The Netherlands, 2018.
30. Levin, E.; Levin, A. Spoken dialog system for real-time data capture. In Proceedings of the Ninth European Conference on Speech Communication and Technology, Lisbon, Portugal, 4–8 September 2005.
31. Black, L.A.; McTear, M.; Black, N.; Harper, R.; Lemon, M. Appraisal of a conversational artefact and its utility in remote patient monitoring. In Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems, Dublin, Ireland, 23–24 June 2005; pp. 506–508.
32. Harper, R.; Nicholl, P.; McTear, M.; Wallace, J.; Black, L.A.; Kearney, P. Automated phone capture of diabetes patients readings with consultant monitoring via the web. In Proceedings of the 15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems, ECBS, Belfast, UK, 31 March–4 April 2008; pp. 219–226.
33. Lucas, G.M.; Rizzo, A.; Gratch, J.; Scherer, S.; Stratou, G.; Boberg, J.; Morency, L.P. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Front. Robot. AI* **2017**, *4*, 51. [[CrossRef](#)]
34. Yokotani, K.; Takagi, G.; Wakashima, K. Advantages of virtual agents over clinical psychologists during comprehensive mental health interviews using a mixed methods design. *Comput. Hum. Behav.* **2018**, *85*, 135–145. [[CrossRef](#)]
35. Ali, T.; Hussain, J.; Amin, M.B.; Hussain, M.; Akhtar, U.; Khan, W.A.; Lee, S.; Kang, B.H.; Hussain, M.; Afzal, M.; et al. The Intelligent Medical Platform: A Novel Dialogue-Based Platform for Health-Care Services. *Computer* **2020**, *53*, 35–45. [[CrossRef](#)]
36. Ireland, D.; Atay, C.; Liddle, J.; Bradford, D.; Lee, H.; Rushin, O.; Mullins, T.; Angus, D.; Wiles, J.; McBride, S.; et al. Hello Harlie: Enabling Speech Monitoring Through Chat-Bot Conversations. *Studi. Health Technol. Inf.* **2016**, *227*, 55–60.
37. Mugoye, K.; Okoyo, H.; Mcoyowo, S. Smart-bot Technology: Conversational Agents Role in Maternal Healthcare Support. In Proceedings of the IEEE 2019 IST-Africa Week Conference (IST-Africa), Nairobi, Kenya, 8–10 May 2019; pp. 1–7.
38. Giorgino, T.; Azzini, I.; Rognoni, C.; Quaglini, S.; Stefanelli, M.; Gretter, R.; Falavigna, D. Automated spoken dialogue system for hypertensive patient home management. *Int. J. Med. Inf.* **2005**, *74*, 159–167. [[CrossRef](#)]
39. Beveridge, M.; Fox, J. Automatic generation of spoken dialogue from medical plans and ontologies. *J. Biom. Inf.* **2006**, *39*, 482–499. [[CrossRef](#)]
40. Clarke, N.L.; Furnell, S.M.; Rodwell, P.M.; Reynolds, P.L. Acceptance of subscriber authentication methods for mobile telephony devices. *Comput. Secur.* **2002**, *21*, 220–228. [[CrossRef](#)]
41. Raza, M.; Iqbal, M.; Sharif, M.; Haider, W. A survey of password attacks and comparative analysis on methods for secure authentication. *World Appl. Sci. J.* **2012**, *19*, 439–444.
42. McDermott, D.S.; Kamerer, J.L.; Birk, A.T. Electronic Health Records: A Literature Review of Cyber Threats and Security Measures. *Int. J. Cyber Res. Educ. (IJCRE)* **2019**, *1*, 42–49. [[CrossRef](#)]
43. Frumento, E. Cybersecurity and the Evolutions of Healthcare: Challenges and Threats Behind Its Evolution. In *m_Health Current and Future Applications*; Springer: Berlin, Germany, 2019; pp. 35–69.
44. Kao, H.C.; Tang, K.F.; Chang, E.Y. Context-aware symptom checking for disease diagnosis using hierarchical reinforcement learning. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LO, USA, 2–7 February 2018.
45. Morreale, S.P.; Spitzberg, B.H.; Barge, J.K. *Human Communication: Motivation, Knowledge, and Skills*; Cengage Learning: Boston, MA, USA, 2007.
46. Glass, J. Challenges for spoken dialogue systems. In *Proceedings of the 1999 IEEE ASRU Workshop*; MIT Laboratory for Computer Science: Cambridge, MA, USA, 1999.

47. Kang, S.; Ko, Y.; Seo, J. A dialogue management system using a corpus-based framework and a dynamic dialogue transition model. *AI Commun.* **2013**, *26*, 145–159. [[CrossRef](#)]
48. Li, Y.; Feng, Z.; Xiao, Y.; Huang, J. A neural network algorithm for signal processing of LFMCW or IFSCW system. In Proceedings of the 1999 Asia Pacific Microwave Conference—APMC'99—Microwaves Enter the 21st Century, Conference Proceedings (Cat. No.99TH8473), Singapore, 30 November–3 December 1999; Volume 3, pp. 900–903.
49. Rasa Documentation. Available online: <https://rasa.com/docs/rasa/> (accessed on 19 March 2020).
50. Unified Medical Language System Documentation. Available online: <https://www.nlm.nih.gov/research/umls/index.html> (accessed on 19 March 2020).
51. Hummer, M.; Groll, S.; Kunz, M.; Fuchs, L.; Pernul, G. Measuring Identity and Access Management Performance—An Expert Survey on Possible Performance Indicators. In Proceedings of the 4th International Conference on Information Systems Security and Privacy, Funchal - Madeira, Portugal, 22–24 January, 2018. Available online: <https://www.scitepress.org/Papers/2018/65577/65577.pdf> (accessed on 23 July 2019)
52. Rehman, U.U.; Lee, S. Natural Language Voice based Authentication Mechanism for Smartphones. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*; ACM: New York, NY, USA, 2019; pp. 600–601.
53. Biswas, M. AI and Bot Basics. In *Beginning AI Bot Frameworks*; Springer: Berlin, Germany, 2018; pp. 1–23.
54. Find Machine Learning Algorithms for Your Data. Available online: <https://mod.rapidminer.com/> (accessed on 22 March 2020).
55. Elliot, A.J.; Maier, M.A. Color psychology: Effects of perceiving color on psychological functioning in humans. *Ann. Rev. Psychol.* **2014**, *65*, 95–120. [[CrossRef](#)]
56. Walker, M.A.; Litman, D.J.; Kamm, C.A.; Abella, A. PARADISE: A framework for evaluating spoken dialogue agents. *arXiv* **1997**, arXiv:cmp-lg/9704004.
57. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, 159–174. Available online: <https://www.jstor.org/stable/pdf/2529310.pdf> (accessed on 23 March 2019) [[CrossRef](#)]
58. Pejovic, V.; Bojanic, S.; Carreras, C.; Nieto-Taladriz, O. Detecting masquerading attack in software and in hardware. In Proceedings of the MELECON 2006—2006 IEEE Mediterranean Electrotechnical Conference, Malaga, Spain, 16–19 May 2006; pp. 836–838.
59. Schrepp, M. User Experience Questionnaire Handbook. In *All you Need to Know to Apply the UEQ Successfully in Your Project*; 2015. Available online: <https://www.ueq-online.org/Material/Handbook.pdf> (accessed on 12 May 2019)



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).