Explainable Retinal Screening with Self-Management Support to Improve Eye-Health of Diabetic Population via Telemedicine

Jannatul Ferdause Tumpa
Marquette University

EXPLAINABLE RETINAL SCREENING WITH SELF-MANAGEMENT SUPPORT TO IMPROVE EYE-HEALTH OF DIABETIC POPULATION VIA TELEMEDICINE

by

Jannatul Ferdause Tumpa, M.S.

A Dissertation Submitted to the Faculty of the
Graduate School, Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

Milwaukee, Wisconsin

December 2021

ABSTRACT

EXPLAINABLE RETINAL SCREENING WITH SELF-MANAGEMENT SUPPORT TO IMPROVE EYE-HEALTH OF DIABETIC POPULATION VIA TELEMEDICINE

Jannatul Ferdause Tumpa, M.S.

Marquette University, 2021

Diabetic Retinopathy (DR) is one major complication of diabetes and is the leading cause of blindness worldwide. Progression of DR and complete vision loss can be prevented by keeping diabetes in control and by early diagnosis through annual eye screenings. However, cost, healthcare disparities, cultural limitations, lack of motivation, etc., are the main barriers against regular screening, especially for a few ethnically and racially minority communities. On the other hand, to well-manage and control diabetes, the diabetic population needs to be physically active and keep their weight healthy. From the perspective of Behavioral Science, Some self-management techniques based on motivational interviewing can be utilized to motivate people to take preventive and mandatory measures to control diabetes. However, technical solutions based on 'Motivational Interviewing' are still not sufficiently available to healthcare providers who work with the diabetic population. Thus, collaborative teamwork of Computer Science and Behavioral Science is contemporary to improve eye health and the overall health of the diabetic population.

In this dissertation, a community telemedicine framework has been proposed and designed which can connect clinicians with community partners to organize retinal screenings in community settings rather than traditional clinical settings. Secondly, automating the initial retinal screenings utilizing Deep Learning models, particularly Convolutional Neural Network (CNN), can reduce ophthalmologists' workload and cost of screening. However, such Machine Learning models lack transparency and cannot explain how these models make particular decisions. Thus, an explainable retinal screening model has been developed to facilitate the recommended annual screening to overcome this limitation. Finally, a Computer-guided Action Planning (CAP) tool has been designed and developed to motivate the diabetic population to adopt healthier behaviors through Brief Action Planning, a self-management support technique. Through several feasibility studies, it is evident that the contribution of this dissertation could be combined to help prevent vision loss from diabetes.

ACKNOWLEDGMENTS

Jannatul Ferdause Tumpa, M.S.

As it's said "It takes a village to raise a child," I have a village of people to thank for contributing to my life and making me the person I am today. I could not have done this without these incredible people by my side.

I want to express my sincere appreciation and heartfelt gratitude to my Ph.D. advisor, Dr. Sheikh Iqbal Ahamed, for his constant guidance and encouragement. Along with research advising, Dr. Ahamed helped me to dream higher in my career path and always made sure that I was giving my best effort in my study and research with his continuous support. I appreciate all the members of Ubicomp Lab for their constant support and motivation, particularly, Zawad, Shiyu, Nihel, Dipranjan, and Riddhi, for their collaboration in different parts of this dissertation.

I am grateful to Dr. Daniel Pinto, who introduced me to some challenging research domains in the intersection of Behavioral Science and Physical Therapy. He made me a part of his BEHTA lab, supported a part of this dissertation research through grants, and mentored me through most of my graduate study at Marquette University.

I am also thankful to Dr. Praveen Madiraju for being on my dissertation committee and for his continuous encouragement and support as faculty and Director of Graduate Studies in the Department of Computer Science at Marquette University. A special thanks go out to other Faculty in the CS department, particularly Dr. Shion Guha, Dr. Nasim YahyaSoltani, Dr. Niharika Jain, Dr. Satish Puri, and Dr. Dennis Brylow, for their teaching and guidance during my course work and TA experience. I want to acknowledge the care and support of all my friends and colleagues in Milwaukee. I will be forever grateful to all the teachers and mentors I have encountered in my life. I gratefully acknowledge that I used the LaTeXtemplate of Paul Kaefer, a Marquette alumnus whom I never met.

I oblige Arthur J. Schmitt Foundation for bestowing me the prestigious fellowship award as an acknowledgment of my leadership contribution and for providing me the financial support during my final year. It helped me to put my entire focus on my research when I needed it most.

I would not be where I am today without the contributions of my parents and siblings, especially my mom, who has been the greatest inspiration of my life. I cannot emphasize enough how grateful I am to my amazing life partner, Riddhiman

Adib, who was my sole inspiration for higher study abroad. He did not let me quit during my weakest moments and has been the constant and unconditional support system throughout my roller-coaster journey. I am also grateful to my parents-in-law for believing in me and always supporting me in all my endeavors.

TABLE OF CONTENTS

ACKNOWLEDGMENTS **i**

LIST OF TABLES **vi**

LIST OF FIGURES **vii**

CHAPTER 1 Preventing Vision Loss from Diabetes **1**

1.1 Diabetes and Vision Loss 1

1.2 Preventing Vision Loss from Diabetes 3

1.3 Community telemedicine to Facilitate Regular Screening 5

1.4 Diabetic Retinopathy and Automated Grading 5

1.5 Managing diabetes through Motivational Interviewing 6

1.6 Organization of this Thesis 8

CHAPTER 2 Design and Development of Community Telemedicine Framework **9**

2.1 Overview of mTOCS system 12

2.2 User Roles, Access Levels and Functionalities 12

2.3 General Acceptance Evaluation 18

2.3.1 Measures 19

2.3.2 Methods 20

2.3.3 Study Results 21

2.4 Screener Software Assessment 23

2.4.1 Measures 24

2.4.2 Methods 24

2.4.3 Study Results 25

CHAPTER 3 Automated Retinal Screening Model with Explainability **27**

3.1 Retinal Screening and its automation 27

3.1.1 Grading Categories 27

3.1.2 Literature Review on Automated Retinal Screening 28

3.1.3 Available Datasets 31

3.2 Explainability and its techniques 31

3.2.1 Post-hoc Global Explainability 33

3.2.2 Post-hoc Local Explainability 35

3.3 Results 35

3.3.1 Training a Convolutional Neural Network 36

3.3.2 Post-hoc Explainability using Grad-CAM 37

3.3.3 Evaluation Metrics for Explainability 38

CHAPTER 4 Diabetes Self-management through Brief Motivational Interviewing 40

4.1 Managing diabetes through Motivational Interviewing 40

4.1.1 Motivational interviewing (MI) 40

4.1.2 Brief MI 41

4.1.3 Brief Action Planning 41

4.1.4 Components of BAP 41

4.1.5 How an electronic version could reduce the training time of MI and make it more available to physicians 48

4.2 Overview of CAP System 49

4.2.1 Basic Features 51

4.2.2 Facilitating Features 56

4.3 Study on Acceptance of CAP among clinicians 58

4.3.1 Study design 59

4.3.2 Data collection 60

4.3.3 Results 61

4.3.4 Discussion 69
4.4 Study on Efficacy of CAP in reducing facilitator training time 70
4.4.1 Study Design 70
4.4.2 Data Collection and Analysis 71
4.4.3 Results 73
4.4.4 Discussion 75
CHAPTER 5 Conclusion and Future Work **76**
5.1 Broader Impact of this dissertation 76
5.2 Future Work 77
BIBLIOGRAPHY **80**

LIST OF TABLES

2.1 Satisfaction survey questionnaire 21

2.2 Quantitative representation of responses for satisfaction survey 22

LIST OF FIGURES

1.1 Fundus retinal image features in Diabetic Retinopathy subjects compared to normal retinal image (Referenced from Vimala 2014) 2

2.1 Work flow in mTOCS 13

2.2 Screener portal: Consent form for participant's signature before screening 14

2.3 Screener portal: Form for collecting demographic information of participants 15

2.4 Screener portal: Form for survey on eye health of participants 16

2.5 Screener portal: The system showing Documentation of Informed Consent 17

2.6 Grader portal: List of features of DR shown, depending on letter type selected 18

2.7 Staff portal: Grading follow-up scenario for a certain participant . . . 19

2.8 Response representation of satisfaction survey questionnaire 22

2.9 Response for question theme "Language", stratified by language preference 23

3.1 Different techniques of achieving explainabilty in Neural Network models [21] 33

3.2 Grad-CAM for retina with and without DR 38

4.1 Flowchart showing the components of Brief Action Planning (citing from Gutnick et. al. 2014) 42

4.2 Flowchart of follow up meeting (citing from Gutnick et al. 2014) . . . 47

4.3 Questionnaire Flowchart in CAP following Brief Action Planning . . 50

4.4 Screenshot of 'Basic Feature: Add new participant' from CAP system 51

4.5 Screenshot of 'Basic Feature: Add new plan' from CAP system 54

4.6 Questionnaire Flowchart of Follow up meeting 55

4.7 Screenshot of CAP system showing the plan versioning feature 56

4.8 An example of prompts guiding CAP interview 57

4.9 Screenshot of CAP system showing the free text-boxes 58

4.10 Screenshot of CAP system showing the Print and Email feature . . . 59

4.11 Age distribution and Proficiency level of clinicians in using technical platforms 62

4.12 Responses for question about facilitation of CAP 64

4.13 Feedback on the prompts shown during interview 67

4.14 Feedback on the free text-boxes on all exit points 67

4.15 Feedback on the Emailing and Printing feature 68

4.16 Feedback on the version support of Action Plans 68

4.17 The skill checklist used to calculate fidelity score 72

4.18 The frequency of checklist questions [‘Green’ means ‘Achieved’, ‘Orange’ means ‘Developing’ and ‘Grey’ means ‘Not applicable’] 73

4.19 Evaluation of student’s recordings by reviewers 74

CHAPTER 1
Preventing Vision Loss from Diabetes

1.1 Diabetes and Vision Loss

Diabetes poses a major threat to human health, and according to the World Health Organization, it is one of the leading causes of death worldwide [48]. About 422 million people worldwide, particularly in low and middle-income countries, currently have diabetes, and this number is increasing exponentially [16]. In the U.S. context, the number of Americans with diabetes is projected to be doubled by 2050 [24]. People who have diabetes for a prolonged period are at higher risk of developing retinal complications that can eventually cause vision loss [64]. In general, Diabetic Eye Disease [8] includes:

- Diabetic Retinopathy: Damage to the blood vessels in the retina.
- Cataract: Clouding of the lens of the eye.
- Glaucoma: Increase in fluid pressure inside the eye, leading to optic nerve damage and loss of vision.

Diabetic Retinopathy (DR) is the most prevalent diabetic eye disease and is one of the leading factors of blindness worldwide [64] and in American adults. It occurs due to changes in retinal blood vessels; in some cases, the blood vessels get

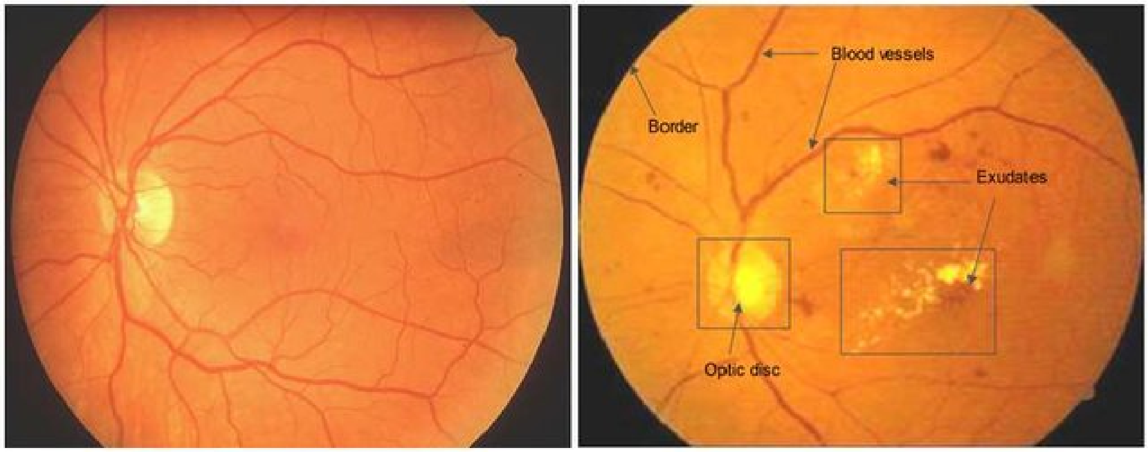


Figure 1.1: Fundus retinal image features in Diabetic Retinopathy subjects compared to normal retinal image (Referenced from Vimala 2014)

swollen and leak fluid. Also, the abnormal growth of new blood vessels on the retinal surface can be responsible for developing DR.

DR does not have any immediately noticeable symptoms, e.g., pain or vision impairment, etc. during the early stages. Vision might get blurred due to Macular Edema, another eye condition which happens when DR affects the macula (the part of the retina responsible for providing sharp and color vision) by swelling due to leak of fluid. Vision is also hampered when new vessels grow on the retinal surface, causing eye bleeding. In some more complicated cases, the disease may progress without any symptoms for an extended period. Due to this type of progression without any symptom, regular eye screening is vital for people with diabetes, especially for those who have diabetes for more than 15 years [14]. A comparison between a normal retina and retina with DR [61] is presented in figure 1.1.

While performing the annual eye examination to screen for DR, in most cases, the eye is dilated to enlarge the pupils to get the fundus image of the retina through a specialized camera. However, this pupil-dilation is a barrier to screening [57] as many patients feel very uncomfortable after this dilation due to the resulting eye-sensitivity afterward. To overcome this issue of discomfort, Scanning Laser Ophthalmoscope (SLO) cameras [23] can be used, which can capture excellent quality retinal images without pupil-dilation.

1.2 Preventing Vision Loss from Diabetes

Multiple racial and ethnic minority communities are disproportionately affected by Diabetic Retinopathy (DR) [9]. Centers for Disease Control and Prevention (CDC) has published several statistical analyses, stating that, Latinos and African Americans are at particularly high risk for Diabetic Retinopathy, and they also have the lowest record of having recommended annual eye exams [26] due to different health disparities prevalent in U.S. [65]. However, Researchers worldwide have concentrated on these health and vision disparities and adopted various approaches to address these challenges [33, 44].

Annual Retinal screening is recommended for diabetic patients, and this imposes a huge workload on ophthalmologists as well as related healthcare workers. Automated grading of diabetic retinopathy has the potential to reduce this workload by supporting the initial screening so that only people with the possibility of eye disease can be referred to ophthalmologists. To maximize the clinical utility of automated grading, an algorithm to detect referable diabetic retinopathy is needed.

Diabetes and related complications are associated with increased body mass index, and regular physical activity (PA) can help maintain healthy body weight and improve the body's ability to use insulin. Despite PA benefits, less than 40% of diabetic population in U.S. [25] engage in regular PA. Thus, there is a need for evidence-based self-management interventions. Patient-centered, individually tailored interventions are associated with improved lifestyle change[55]. One example of a patient-centered intervention is Motivational Intervention (MI). However, this communication process is still not available to integrate in regular clinical flow due to technical constraints.

In summary, DR cannot be prevented; however, the risk can be significantly reduced through managing diabetes and facilitating regular retinal screening. After conducting a comprehensive literature survey[55, 30], further described in chapter 2, chapter 3 and chapter 4, we summarised the following steps that can lead to improved healthcare infrastructure to support the diabetic population to take better care of their eye health:

- We can develop a community telemedicine framework to enable the community partners to arrange regular eye-screening sessions during community events.

- We can develop an automated retinal screening model to reduce eye doctors' workload and provide regular screening to more people.
- We can develop an action planning tool based on the concepts of Motivational Interviewing, which can facilitate doctors to make plans with diabetic patients about managing their diabetes.

1.3 Community telemedicine to Facilitate Regular Screening

To address the issue of healthcare disparity among minority communities, and to provide comfortable, fast, and cost-effective eye-care for diabetic people, a collaborative telemedicine approach in community locations during regular hours and community events can be greatly beneficial. This has been further discussed in chapter 2. On that note, although there are many software systems like Epic [22], Cerner [13] to provide Electronic Health Record services for clinical care, there are no established systems available to support the collaborative communication among community partners for telemedicine in community settings.

1.4 Diabetic Retinopathy and Automated Grading

Extensive research has been conducted about automating the initial screening of DR [30, 18, 10, 60, 63]. Literature shows that a convolutional neural network (CNN) performs with accuracy up to 96% [63] for grading of retinal images. Although the AI-based models have achieved high accuracy in the diagnosis of DR, the current challenge is the clinical validation and real time deployment of these models in clinical practice. Most of the studies used training sets from a homogenous population of a region or a publicly available dataset. Diversifying the dataset in terms of ethnicities and cameras to capture the images will somewhat address this challenge. The other challenge is the 'black- box' phenomenon. In Neural Network models, it is challenging to understand how exactly a model reaches a particular decision, or to identify which exact features it utilizes. How can the results of AI-based algorithms be properly understood by clinicians and researchers? How can we ensure the reliability of algorithms if we cannot understand how they operate? Interpretable machine learning would be an effective tool to mitigate these problems. It gives machine learning models the ability to explain or to present their behaviors in understandable terms to humans, which is called interpretability or explainability.

1.5 Managing diabetes through Motivational Interviewing

Soderlund (2017) [55] has conducted a review to examine Motivational Interviewing's effectiveness for physical activity self-management for adults diagnosed with type-2 diabetes. In the review article, Soderlund has analyzed nine studies to assess how motivational interviewing methods have been applied to physical activity interventions for adults with diabetes mellitus type and what motivational interviewing approaches are associated with successful physical activity outcomes with diabetes mellitus 2. Findings suggest that MI can effectively improve physical activity self-management for the diabetic population and also identified three important key points about using MI:

- MI sessions should target a minimal number of self-management behaviors, i.e., instead of targeting multiple behavior change plans altogether, it is more effective to target one or two particular behaviors to have a better outcome.
- Either frequency or duration of MI sessions should be emphasized. If the sessions are frequent, the duration should be kept minimal, and vice versa.
- The facilitators of MI should be proficient with the spirit of MI, have proper training, and focus on the patient's autonomy while sketching the behavioral change plan.

Gutnick et. al [31] introduced "Brief Action Planning" (BAP), a form of brief MI, which is well-structured, based on principles of MI, and can be quickly adopted in clinical care. However, there is no technical support or electronic tool to integrate BAP or similar methodologies in regular clinical flow. Thus, developing an action planning tool based on the concepts of Motivational Interviewing, which can facilitate doctors to make plans with diabetic patients about self-management can be a contemporary and novel contribution in healthcare research.

1.6 Organization of this Thesis

The remainder of this thesis expands upon the research challenges identified in this chapter. In Chapter 2, we will discuss about the development of a community telemedicine framework which can connect the community partners with eye doctors to arrange regular eye-screening sessions during community events. Chapter 3 will explain different techniques of achieving interpretability in Neural Network models, the development of automated retinal grading model utilizing retinal images and finally, the level of explainability we have achieved utilizing Grad-CAM method. Design and development of an action planning tool based on concept of 'Brief Action Planning', which has been found to be effective in diabetes self-management, will be discussed in Chapter 4, along with evaluation results of the tool among different cohorts. Finally, we will conclude in Chapter 5 with ideas for future research that are inspired by our work.

CHAPTER 2
Design and Development of Community Telemedicine Framework

To overcome the existing vision health disparities [68] among socioeconomically disadvantaged groups, a paradigm shift through innovative interventions is necessary. Community telemedicine refers to a collaborative approach through creating a multi-sector research team representing distinct areas of expertise where each individual serves a specific role in building a strong and effective partnership. This approach can address health disparities and improve eye health by focusing on individual communities [33]. Hispanic/Latinos are identified as a high-risk population for diabetes [44] and diabetic eye diseases. They also face additional constraints resulting from socioeconomic, cultural challenges, and a lack in the number of Spanish-speaking eye-specialists [20, 49]. Compared to English-speaking, non-Latino counterparts, non-English-speaking Latinos reports 22% fewer visits to physicians in the state of Wisconsin, USA [46]. Therefore, to address vision health disparities, a collaborative telemedicine approach can facilitate the improvement of eye-health among some specifically disadvantaged communities by providing special attention to their common barriers.

Availability of internet and IT tools, access to large-scale datasets [37], improvement in diagnosis accuracy by novel machine learning models [15], artificially intelligent systems [36], and finally, the growth of digital technology, in general, has led recent research to be focused on development of mobile health (mHealth) tools to aid in healthcare systems. Specialized mHealth tools [12] have been built for the treatment of DR; however, they have been focusing on traditional clinical solutions with minimal focus on human-computer interactions. Access to affordable healthcare for all is a large crisis [27], and existing tools fail to address issues that possibly could mitigate it. One of the existing mHealth solutions, EyePACS [19], is focused on improving retinal care for diabetic patients, and has been used by clinicians to provide a fast diagnosis. However, our proposed framework focuses more on the active roles of community partners as a facilitator of connection and communication with specific communities. In addition to that, considering the barrier of healthcare insurance cost [32], our framework also provides such care outside the hospital in a community event, where more people with no health insurance can receive eye-health checkups. Finally, a fully customized Electronic Health Record (EHR) system is needed to support community groups.

With the aim to bridge this gap, we have designed and developed a software framework named, “Mobile Teleophthalmology in Community Settings (mTOCS)”. We have trained several bilingual staff from United Community Center to use our system and perform eye screening events at Hispanic/Latino community events. Additionally, our system stores and provides fundus retinal images to eye-specialists, captures online diagnosis, and follows up on those who were found to have retinopathy to assess best referral modality and further barriers to care.

In this project, ‘participants’ refer to general people seeking retinal screening at a community event, ‘screeners’ refer to staffs from community partners (e.g., United Community Center, Milwaukee, USA) using our software and managing the complete screening process (data entry, DR image collection and storage, etc.), ‘graders’ refer to specialized physicians (e.g. ophthalmologists) who diagnose and grade the retinal images, and ‘staffs’ refer to general support staffs (e.g. Milwaukee Health Department) who receive the grading and sends out relevant letters with diagnosis and recommendations to the participants. Difference between screeners and staffs is that, screeners actively work as community support whereas staffs work behind the curtain and maintain official tasks as needed. We present the complete workflow in Figure 2.1.

Our research work establishes and evaluates a novel approach of providing teleophthalmology support in community settings for improving eye-health among a specific community (in our research, Latino/Hispanic). By conducting a feasibility study among 400 participants from our community screening events, we have found that this approach can significantly benefit this community. Especially, the bilingual/native staffs, using our multilingual mTOCS framework can help address language and culture barriers. The screener assessment study reported in **??** has shown that the mTOCS framework can improve efficiency by saving time and providing an intuitive and usable service.

2.1 Overview of mTOCS system

Our mTOCS software framework has been designed with a focus on simplicity, efficiency and efficacy on promoting the collaboration between multiple moving agents (screeners, graders, participants, etc.). This section describes the design methodology of the software framework along with the different roles with their actions and access levels, and the features that makes it beneficial for conducting eye-screening events in community settings.

2.2 User Roles, Access Levels and Functionalities

The unique strength of our mTOCS software framework is its availability and support in community settings. One of the goals of "Tele-eye Health Project" is not to bring people to a clinical setting for a service; rather, we brought our service to the community in their comfort zone. Free eye-screening events were provided offline through word-of-mouth, fliers, social events as well as online by posting on social media and advertisements with the support from the City of Milwaukee Health Department. The screening events were held in community centers, health fares, food panties, schools, Health Department clinics, etc. Our community partners conducted screening events covering the most convenient community

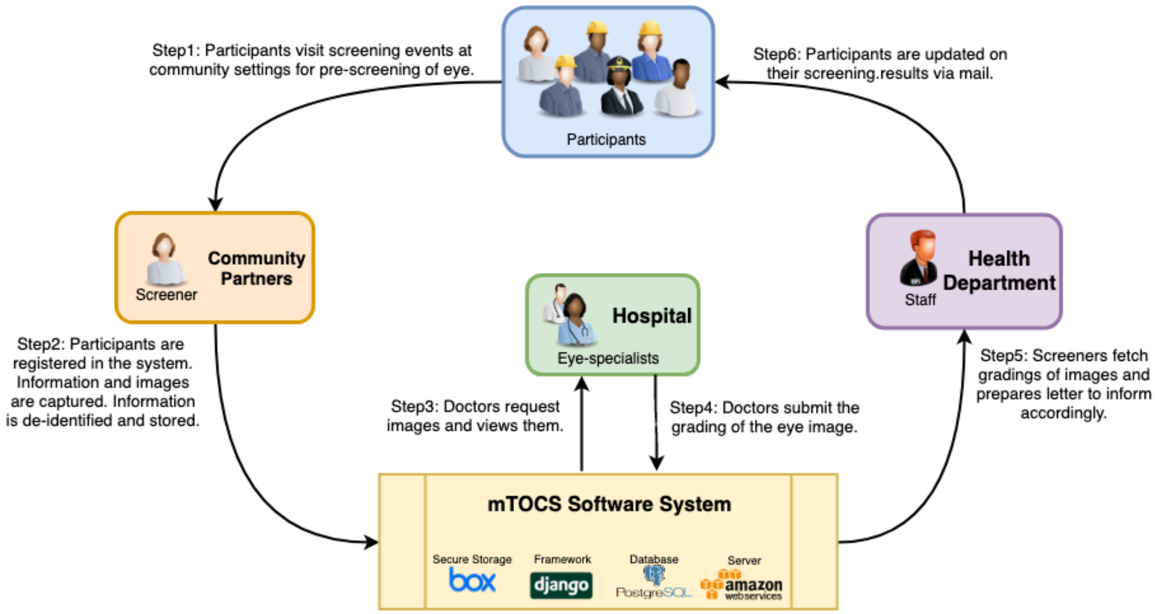


Figure 2.1: Work flow in mTOCS

locations and each screening event continued for 3 or 4 hours, covering 18-20 participants in each event.

A general scenario of an eye-screening goes as follows: participants were greeted at the reception and added to a queue where they can wait for their turn to be screened. The event locations had arrangements of adequate seating space in a comfortable waiting area. When a participant's turn comes up, they were called and seated in a separate area. The screeners greeted them and shared the project's general goals and steps. The participants were provided with a consent form in an electronic media where they read through the terms and conditions and provided their signature (Figure 2.2).

Upon receiving consent, their demographic information was collected and

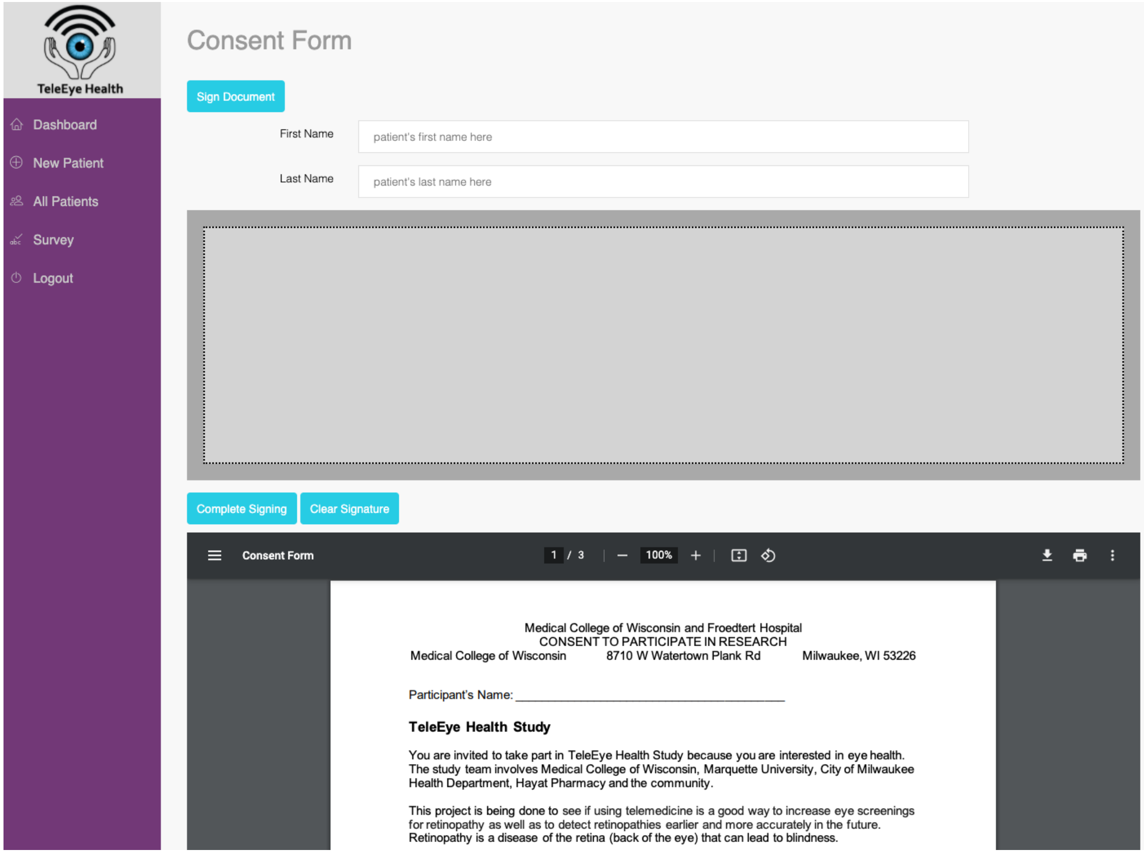


Figure 2.2: Screener portal: Consent form for participant's signature before screening

stored using our mTOCS software as shown in Figure 2.3. The next step was a survey of eye-health as shown in Figure 2.4 which was utilized by the ophthalmologists while grading the retinal images. The survey and discussion were completed in the preferred language of the participant, in which he/she was most comfortable and fluent. The screening tool shows a Documentation of Informed Consent at the end of this survey where the screener marks whether the participants meet all the criteria required to be included the specific study and how the participants want to receive a copy of their consent form. The flow of this event is

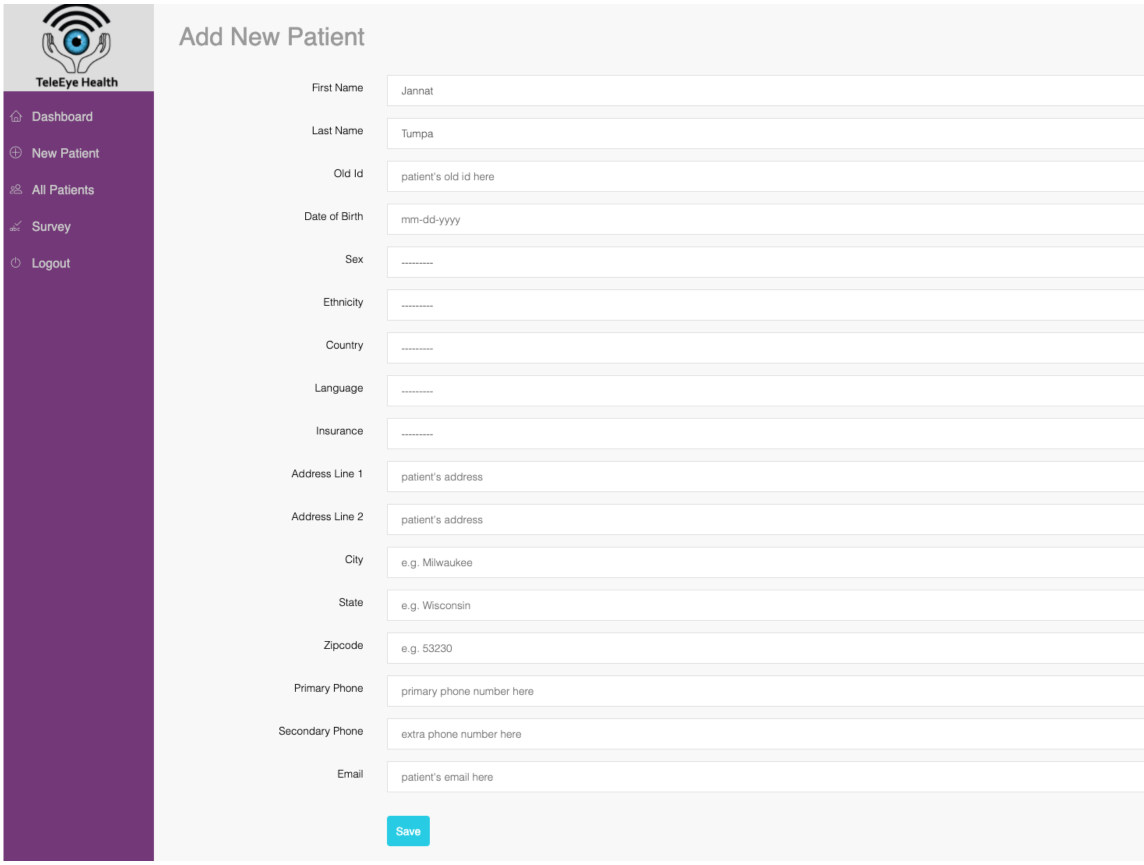


Figure 2.3: Screener portal: Form for collecting demographic information of participants

shown in Figure 2.5 and if the ‘email’ field is selected, an automated email is sent to the participant with the consent form as attachment.

Participants were then requested to move to a separate dark room with SLO fundus camera set up, where images were taken using the SLO fundus camera. The SLO fundus camera used in our screening events captures non-mydriatic fundus images so that participants visibility is not obstructed for the next few hours. After capturing the fundus images, participants were informed about the time when

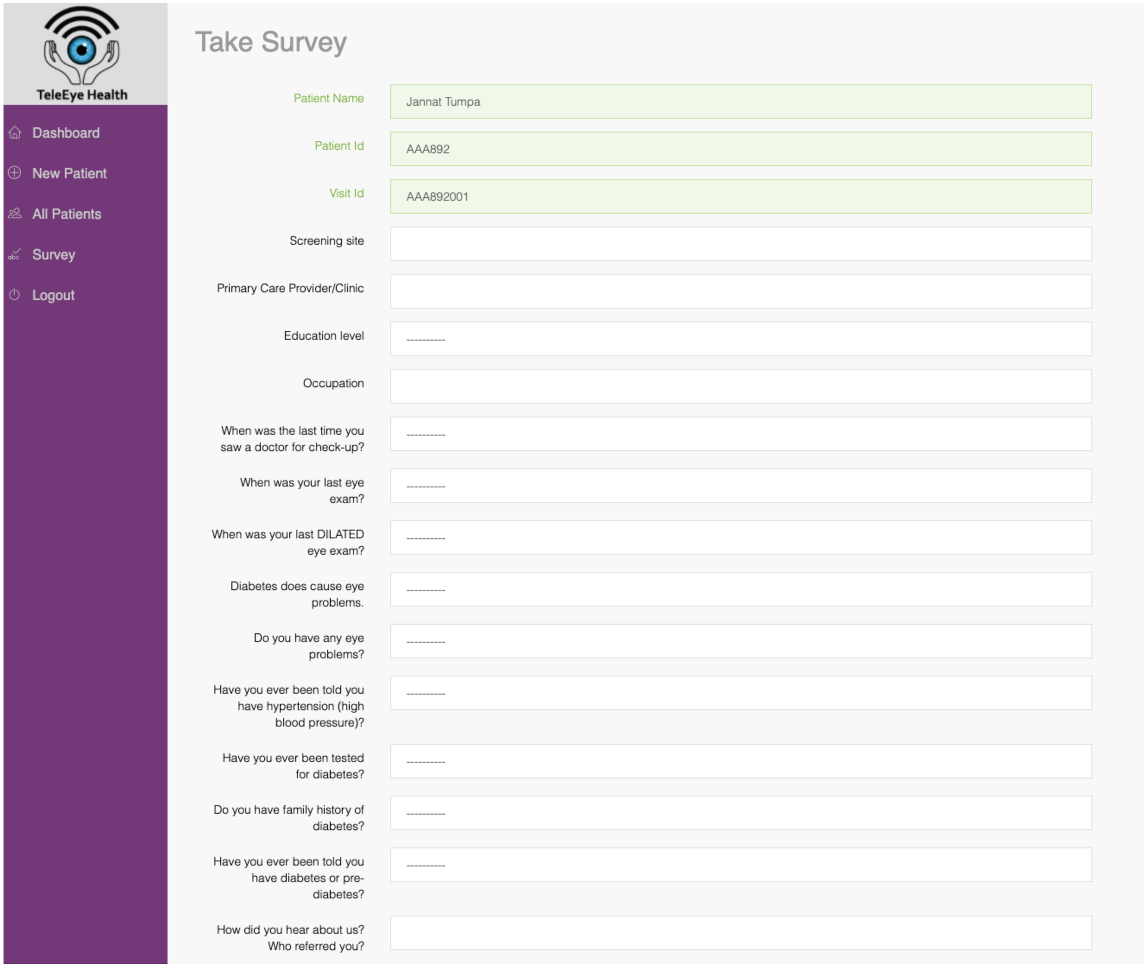


Figure 2.4: Screener portal: Form for survey on eye health of participants

he/she should expect screening results along with the future updates of screening events.

Depending on the responsibilities performed by the community partners and eye-specialists involved in 'Tele-eye Health Project,' the mTOCS framework has four different access roles with distinguishing features and support:

1. **Screener Portal *(Only accessed by screeners)*:** Participant registration

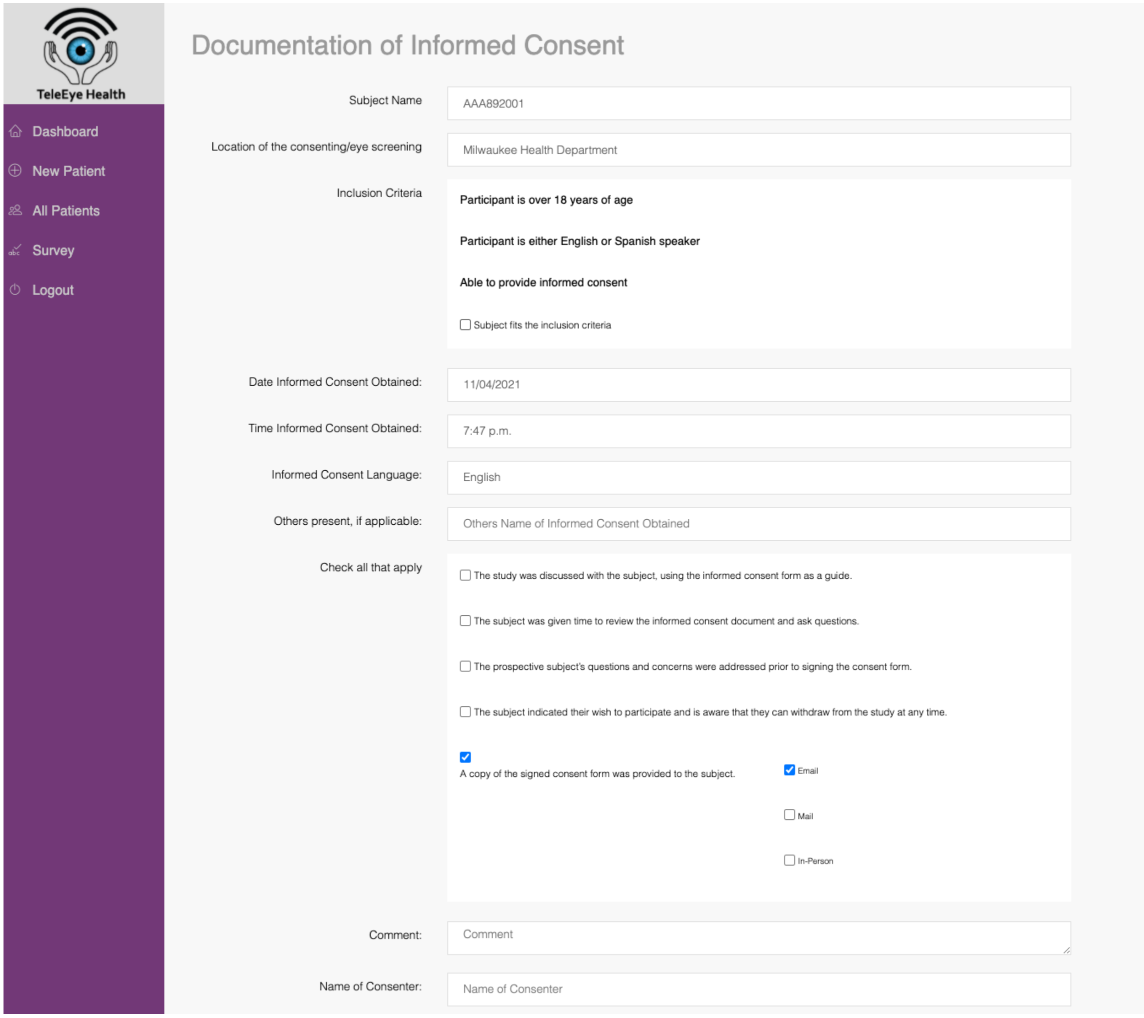


Figure 2.5: Screener portal: The system showing Documentation of Informed Consent

(with demographic information), Language preference, Survey completion, DR image capture and storage

2. **Grader portal** ***(Only accessed by graders)*****:** View ungraded image, Grade images, Change/update graded images

3. **Report Distribution Portal** ***(Only accessed by staffs)*****:** View retinal

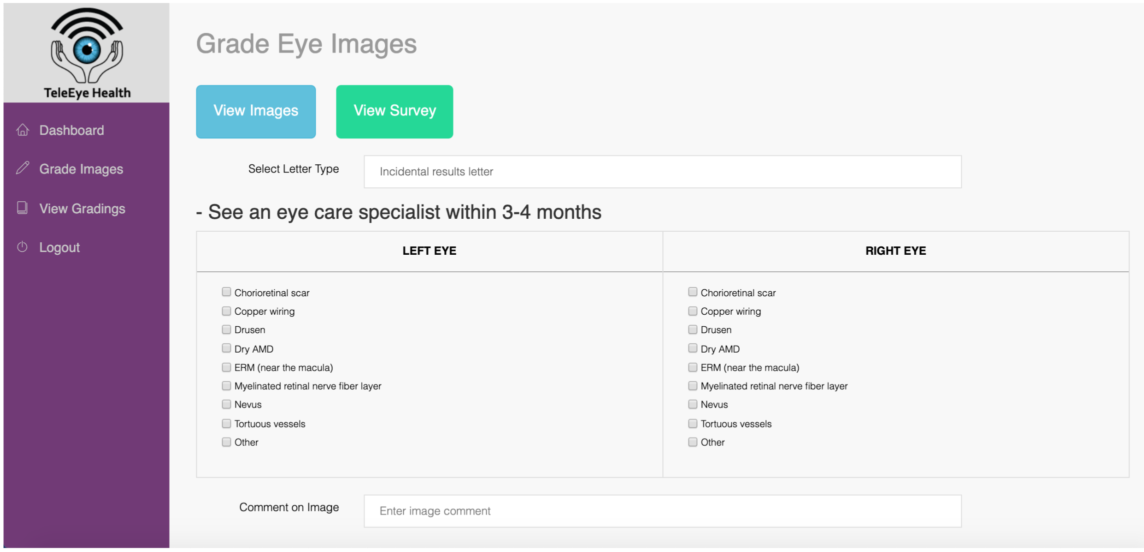


Figure 2.6: Grader portal: List of features of DR shown, depending on letter type selected

image grading, Easy printout of preformatted letters with image grading, Future communication, Request to follow-up

4. **Data management Portal *(Only accessed by system admin)*:** Data export/ backup procedure, Data analysis

2.3 General Acceptance Evaluation

Our tele-eye health project, mTOCS, has multiple dimensions of user interactions; on one end, it has to perform fluently as part of an interview, on the other hand, it serves as an intuitive tool to be used by the screeners without minimal disruption. To capture the strengths of these dimensions and to evaluate the impact of our proposed software framework, we developed and conducted a participant

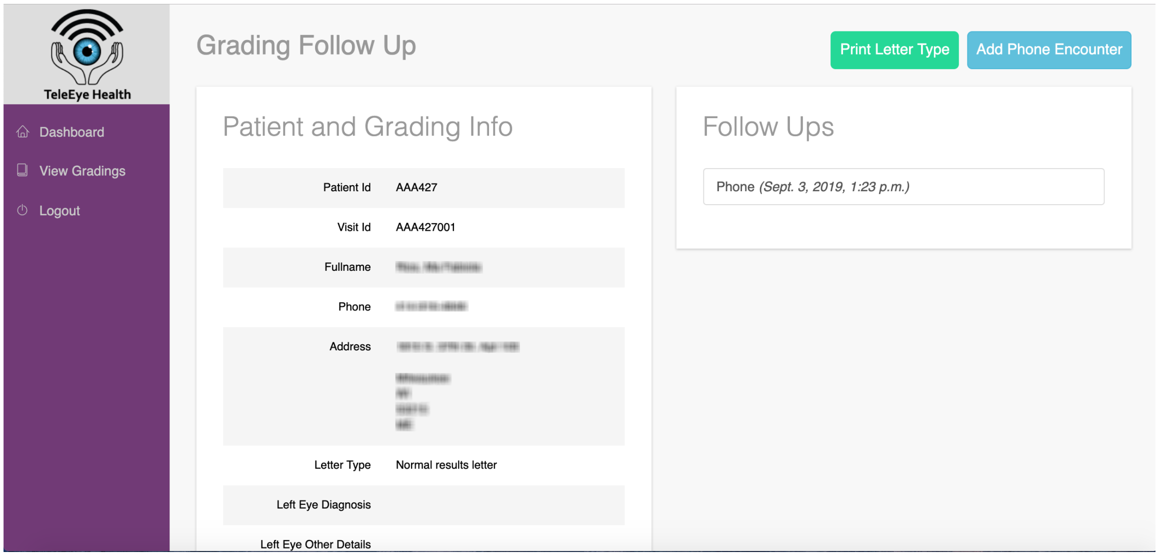


Figure 2.7: Staff portal: Grading follow-up scenario for a certain participant

satisfaction survey, addressing the comfort and convenience of community members in a tele-eye health project.

In the participant satisfaction survey, through collecting general participant feedback, we wanted to explore whether the participants found public eye-health screening acceptible. Our goal was 1) to assess the acceptance level and attitudes of participants regarding teleophthalmology in community health settings in Milwaukee, 2) to analyze the strengths and weaknesses of the eye-screening events through comparing likes and dislikes among participants.

2.3.1 Measures

The survey contained a total of eleven(11) questions. Eight(8) of them are 5-point Likert scale questions, addressing various aspects of the community setting eye-screening which includes comfort, location, privacy, involvement, dissemination, connection, language, and general acceptance. Options of responses ranged from 1 (strongly disagree) to 5 (strongly agree). Details of each statement and the corresponding theme are presented in table 2.1. The other three(3) are open questions about what the participants liked about the screening, disliked about it and whether they would discuss it (or recommend it) to their friends and families. The survey does not directly focus only on the software, instead discusses issues that connect the framework as a whole. This involves the whole ecosystem consisting of location, timing, use of camera hardware-software combination, and screening procedure in general.

The participants were asked to complete the satisfaction survey after they completed their free eye-screening. Although self-reported survey responses are not always the best representation of the entire scenario, they can convey critical information and valuable feedback from a different perspective, and points towards a general acceptance of the targeted community .

2.3.2 Methods

A total of 400 satisfaction surveys were given to participants after completing the retinal screening process in either English or Spanish. Excluding incomplete or no responses, we collected a total of 378 satisfaction surveys. The surveys were provided to the participants at the end of the respective eye-screening and accumulated through multiple eye-screening events sessions. The results presented

Statement	**Theme**
I was comfortable during the telemedicine session	Comfort
It was convenient to get eye screening at this location	Location
I was worried about my privacy using this type of eye screening	Privacy
This type of screening helped me get more involved with my health	Involvement
I would recommend this type of screening to others	Dissemination
I liked seeing the image of my retina	Connection
It helped having bilingual/ Spanish staff do the screening	Language
I would use this type of screening again	General acceptance

Table 2.1: Satisfaction survey questionnaire

in the following section shows that, on average, the participants were comfortable throughout the screening process.

2.3.3 Study Results

The response rate for our satisfaction survey was 94.5% (378 out of 400). We had a total of 225 (59.52%) participants preferring English for communication, 103 (27.25%) participants preferring Spanish for communication, and 50 (13.23%) participants with no language preference. The participants with no preference indicate either being comfortable in both languages, having no preference over language or unwilling to submit a response to this specific question.

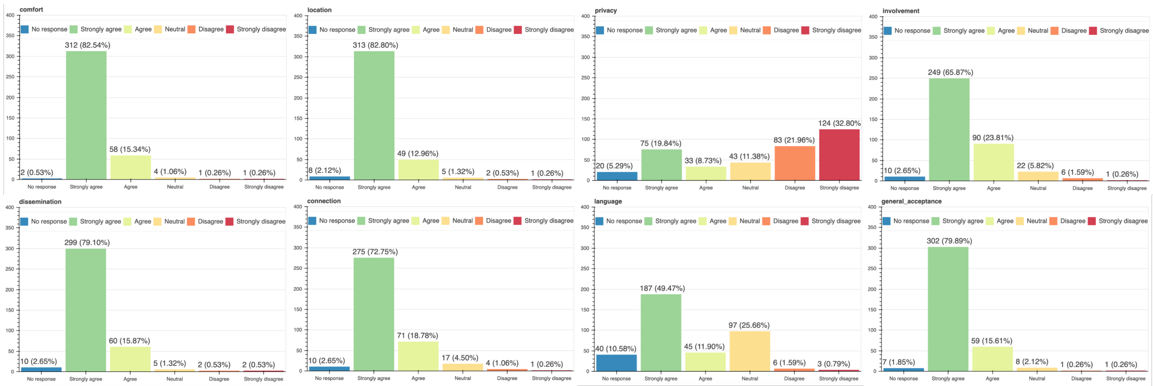


Figure 2.8: Response representation of satisfaction survey questionnaire

In general, the responses from the participants were very positive. The responses for Likert-scale questions are represented in graphs in the Figure 2.8. Calculating based on a scoring of 1 for a response of ”strongly disagree” to 5 for ”strongly agree,” a mean (average) score of more than 3 signifies a general positive response towards the specific question. The number of responses and their mean, standard deviation for individual questions is described in Table 2.2.

Theme	**Count**	**Mean**	**Standard Deviation**
Comfort	376	4.81	0.48
Location	370	4.81	0.49
Privacy	358	2.58	1.54
Involvement	368	4.58	0.70
Dissemination	368	4.77	0.55
Connection	368	4.67	0.64
Language	338	4.20	0.98
General acceptance	371	4.78	0.51

Table 2.2: Quantitative representation of responses for satisfaction survey

Based on the average mean of a score higher than 4, we find that, in general,

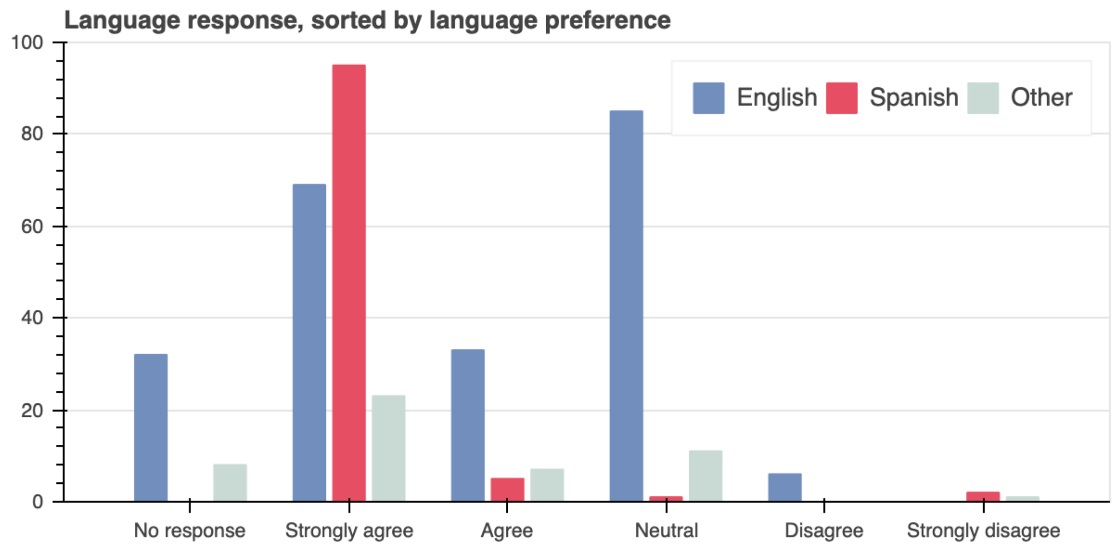


Figure 2.9: Response for question theme "Language", stratified by language preference

the community eye-screening event satisfies most of our participants. The question about privacy and language does not have a high positive response, as seen in other questions. Unlike other questions, the question regarding privacy (*"I was worried about my privacy using this"*) is negatively constructed. Because of this, a low score (2.58, which is less than 3) actually signifies that the participants are not much worried about their privacy. However, we also find a difference of opinion on that in between participants.

2.4 Screener Software Assessment

Compared to the number of participants, we have a limited number of human-resources who act as screeners using the mTOCS software in the eye-screening events. To capture their viewpoint, we conducted a separate survey focusing on general software assessment, focusing on four different aspects of our software system. The four focus points for evaluation of our system from the viewpoint of screeners are (1) feasibility, (2) usability, (3) viability, and, (4) satisfaction. This feedback can add value to our system structure and project continuation, since the screeners have undergone multiple iterations of the mTOCS software.

2.4.1 Measures

The survey was designed starting with a general introductory section. It consisted of demographic questions regarding the screeners participating the survey. This section collected information on age group (18-24, 25-34, 35-44, 45-54, and 55+) and the survey-participating screeners' technical proficiency in using new softwares, in a Likert Scale question. Age and technical proficiency are generally found to be correlated [38, 29] and are both potential indicators of successful field usage of our software system.

2.4.2 Methods

Although our software system was being used to serve more than 1400 people from community, access to it is limited to a handful of trained screeners. We conducted our survey for screener software assessment to only those who actively used mTOCS system to collect data. A total of five (5) complete responses were collected from five individual screeners. The screeners gave consent to collect their responses and in general it took 15-20 minutes to complete the complete set of questionnaire.

2.4.3 Study Results

The survey was conducted among five (5) screeners who actively used mTOCS system and the study result was very straightforward and overall positive. We addressed the actual response they agreed to within parentheses in the following parts.

The screeners were of diverse age group (1 in age 18-24 group, 2 in age 25-34 group, 1 in age 35-44 group, and 1 in age 45-54 group). Two of them had strong technical proficiency *(I figure out all functionalities with little effort all by myself)*, two had fair technical proficiency *(It takes effort, but I get all the features by myself)* and one had average technical proficiency *(I can use basic features, but often need training for advanced features)*.

Under feasibility, 4 out of 5 (80%) mentioned the software saved a lot of time. On average, completing the survey with paper took 15.5 minutes whereas the software took only 6.5 minutes, which is less than half of paper survey. 80% rated the software highly efficient compared to the paper form.

Under usability, 3 out of 5 (60%) of screener survey participants marked our mTOCS software system as highly usable *(I could use all functionalities with ease)* as well as highly intuitive *(I could figure out all functionalities very easily without any help)*. Even though, all of them needed a little training using the software *(I needed training to figure out the basic features)*.

Under viability and satisfaction, 3 out of 5 (60%) mentioned our software still contained errors *(The software runs smoothly most of the time, with a few hiccups)*, and the other 2 (40%) mentioned facing lesser number of errors *(We find once a few error here and there)*. 4 out of 5 (80%) were also highly satisfied with the mTOCS software, based on its user interface, features and in general, performance, and the other 1 (20%) were somewhat satisfied.

CHAPTER 3
Automated Retinal Screening Model with Explainability

3.1 Retinal Screening and its automation

3.1.1 Grading Categories

While summarizing the literature work in subsection 3.1.2, we have found that different studies have used different grading categories, for example, some AI systems have used binary information like presence or absence of DR, some have classified images into either Referable Diabetic Retinopathy or non-Referable DR. Referable diabetic retinopathy, also known as sight-threatening DR (STDR) is defined as any retinopathy more severe than mild diabetic retinopathy, with or without diabetic macular edema. Recent work using multi-staging or ICDR grades to classify the disease. The binary classification is sufficient for screening purposes and to refer the patients when to see an ophthalmologist; however, the five-point gradings provide information about the referral's urgency.

According to the American Academy of Ophthalmology, there are five Diabetic Retinopathy levels based on the findings observable on fundus images, which is also known as the five-point International Classification of Diabetic Retinopathy (ICDR). The five grades are:

1. **No Apparent Retinopathy:** No abnormalities

2. **Mild Non-proliferative Diabetic Retinopathy (NPDR):**

   Microaneurysms only

3. **Moderate NPDR:** More than just microaneurysms but less than severe NPDR

4. **Sever NPDR:** More than 20 intraretinal hemorrhages in each of 4 quadrants and no sign of Proliferative retinopathy.

5. **Proliferative DR:** Neovasculaization or vitreous hemorrhage

3.1.2 Literature Review on Automated Retinal Screening

One of the earliest studies on automatic detection of DR from color fundus photographs was by Abramoff et al. in 2008 [5]. It was a retrospective analysis done with non-mydriatic images from the EyeCheck DR screening project. They were able to detect RDR with 84% sensitivity and 64% specificity. In 2013, Abramoff et al. [4] published the sensitivity and specificity of the Iowa Detection Program (IDP) to detect RDR and found a high sensitivity of 96.8% and specificity of 59.4%. The area under the AUC was 0.937. In 2015, Solanki et al. [56] used their EyeArt AI software with the Messidor2 dataset. EyeArt screening sensitivity for RDR was 93.8%, the specificity was 72.2%, and the AUC was 0.94. Since 2012, a large number of commercially available software systems were developed for the automated detection of DR, known as automated retinal image analysis systems

(ARIAS). Tufail et al. [59] conducted a study in 2013 and published in 2017 to evaluate these systems. Retinal images analyzed by three automated retinal image analysis systems, namely iGradingM (UK), Retmarker (Portugal), and EyeArt (USA) were compared to standard manual grading of DR by human graders/ophthalmologists. EyeArt and Retmarker have higher sensitivity for RDR than human graders. Abramoff et al. [6] in 2016 showed in their study that the integration of CNN to an existing DR detection algorithm resulted in improved detection of RDR by minimizing the number of false positives. Using the Messidor-2 dataset, a sensitivity of 96.8%, and specificity of 87% for RDR was obtained in their study. The specificity improved from 59.4 to 87% when compared with their previous study in 2013. This hybrid screening algorithm, known as the IDx-DR, became the first commercially available AI device to get US Food Drug Administration (FDA) approval for DR screening in April 2018. The IDx-DR can detect RDR (more than mild [mtm] DR) with a sensitivity of 87.4% and a specificity of 89.5%. In 2016, Gulshan et al. [30] reported the Google DL DR validation study results. The algorithm was trained using 128,175 macula-centered retinal images obtained from EyePACS in the United States and retinal images from eye hospitals in India. In the break-through major validation study of the Google algorithm for DR detection, Gulshan et al. reported high sensitivity and specificity for RDR (sensitivity of 97.5% and specificity of 93.4% in the EYEPACS-1 and 96.1% sensitivity and 93.9% specificity for Messidor-2 set). A study by Gargeya et al. [28] with another DL algorithm to detect all stages of DR, showed a sensitivity of 94% and a specificity of 98% for RDR, with an AUC of 0.97 with EyePACS. External validation was done on the MESSIDOR-2 and E-Ophtha datasets in this study. Their study focused on the identification of mild NPDR and not just RDR.

The most recent major study reported on validation of DL was by Ting et al. [58] in Singapore. Their study included multiple retinal images taken with conventional fundus cameras from multiethnic cohorts of people with diabetes. Their algorithm showed high sensitivity and specificity for identifying DR and other eye diseases such as age-related macular degeneration and glaucoma. The sensitivity and specificity for RDR was 90.5% and 91.6%, respectively, and for STDR, the sensitivity was 100%, and the specificity was 91.1% in their study.

In other reported DR screening algorithms, the sensitivity varied from 87 to 97%, and the specificity from 59 to 98% . So a majority of the available AI methods would be capable of being used for DR screening according to the requested FDA endpoints. Most of them seem to be performing better and faster than clinicians. However, a direct comparison of different algorithms is difficult because of significant differences in grading rubric, grader experience, image quality, and reference standard. However, the majority of the works have mentioned the lack of interpretability as their limitation. The ability to explain the result of a model is an integral part of decision making in the health sector.

3.1.3 Available Datasets

There are many publicly available annotated datasets of retinal images which have different goals, characteristics and labeling. Three widely used public datasets which are labeled using the same five-point ICDR gradings are:

- Diabetic Retinopathy Detection [2]: Around 44000 images
- APTOS 2019 Blindness Detection [3]: Around 4500 images
- Messidor-2 [1] : 1748 images

Any or all of these datasets can be utilized in training a Machine Learning model for automated grading of retinal images. The advantage of using a public dataset is that these datasets are already preprocessed and ready for use in model training. However, since retinal images can significantly vary depending on camera used, population, lighting environment, etc., the model trained on any of these datasets might not perform great in any real-life scenario. Thus, we can use these datasets for pretraining and fine-tune the model using the real-time images suited for any particular application.

3.2 Explainability and its techniques

Interpretability and Explainability are two interchangeably used terms with closely related meanings. Interpretability indicates what is happening in the model in technical terms, whereas explainability helps to understand why it is happening in human terms, and both terms are used interchangeably. Du et al. [21] have summarised the different ways to achieve interpretability in Machine Learning models and have grouped the techniques into two broad categories:

- **Intrinsic Explainability:** It is achieved by constructing self-explanatory models that incorporate explainability directly to their structures.
- **Post-hoc Explainability:** It provides explanations for an already developed model.

The main difference between these two groups lies in the trade-off between model accuracy and explanation fidelity. The post-hoc technique can be again be categorized into global and local methodologies and also have many sub-categories.

In discussing post-hoc local explainability [21] in machine learning models, we focus on the local behavior of the model, which means we look for identifying contributions of each feature in the input toward a specific model prediction. Broadly there are two different approaches here: model-agnostic approach and model-specific approach. As defined, the model-agnostic approach tries to explain the predictions of machine learning models irrespective of the implementation. In contrast, model-specific approaches treat the models as white boxes and extensively use internal structure to explain.

Model-specific methods are distributed into three separate and broad

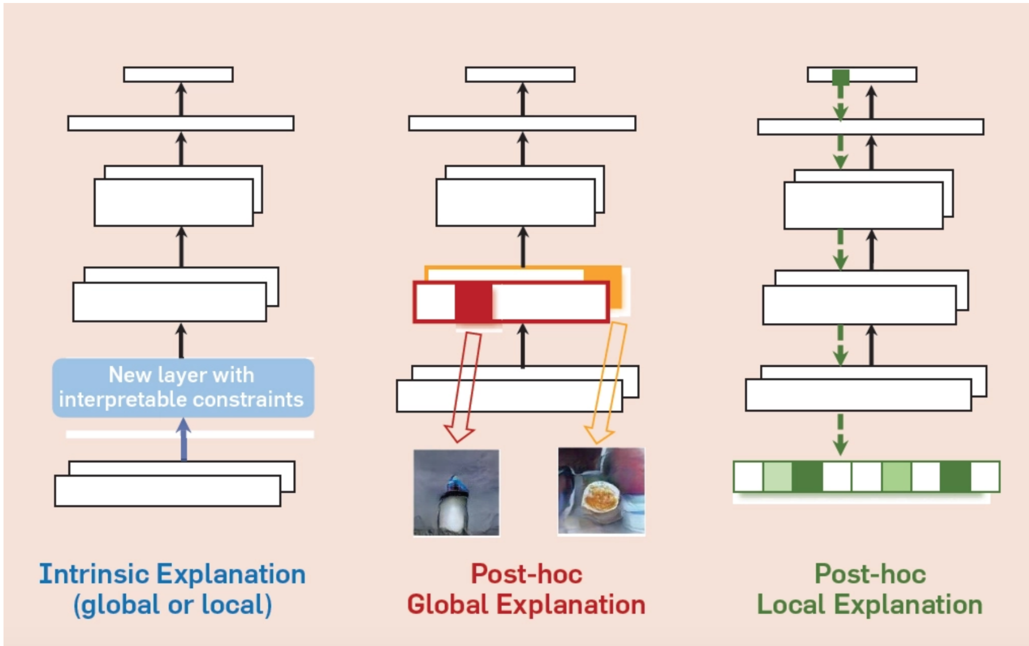


Figure 3.1: Different techniques of achieving explainabilty in Neural Network models [21]

categories: (a) back-propagation based explanation, (b) perturbation based explanation, and (c) investigation of deep representations.

After reviewing all the techniques from review paper of Du et al.[21], we have found that three of them have some possibilities to provide interpretability in DR assessment, which are Intrinsic, Local Post-hoc, and Global Post-hoc.

3.2.1 Post-hoc Global Explainability

Post-hoc global interpretation strategy tries to explain the global understanding of a trained model [21]. As for CNN, the learned knowledge is not human representable; the explanation aims to understand the representations captured by the neurons.

One widely used CNN representation strategy is to find the preferred input that activates a neuron [45] and visualize this information. Several visualization techniques are based on computing the gradient of the class score for the input image. The first one generates an image, which maximizes the class score, thus visualizing its notion. The second technique computes a class saliency map, specific to a given image and class. We will use our pretrained model generated by DenseNet-121. Our goal from this global post-hoc technique is to find specific regions or information that the model uses to detect diabetic retinopathy.

Several visualization techniques can be categorized as follows.

Visualizing Convolution Layers

In this technique [66], the intermediate activations are visualized for each layer of an input image. The initial layers should look for the retina's basic shapes, and the ending layers should be interested in more detailed information about diabetic retinopathy.

Visualizing Class Activation Map (Grad-CAM)

In general, the Class Activation Map (CAM) [69] produces a map from CNNs where global-average-pooled feature maps are fed into a softmax and a gradient is used. In short, this procedure finds a penultimate layer for which its gradient should also be in consideration. There is a general version of CAM named Grad-CAM [54], which uses a weighted gradient from that penultimate layer.

3.2.2 Post-hoc Local Explainability

One of the most noteworthy algorithms in model-agnostic, post-hoc local explainable methods is Local Interpretable Model-agnostic Explanations (LIME) [52]. LIME aims to generate an interpretable model from the interpretable representation, which is locally faithful to the primary classifier. In simpler terms, after an ML model is built for prediction, LIME uses both the main dataset and predictions and leverages those to report which features contributed most to a single prediction. It starts with detecting "super-pixels" from an image *(cluster of pixels carrying similar information, or contributing equally to a specific prediction)*, and then goes through iterations of perturbations in them in a specific image. The goal is to detect which super-pixels contribute most to a specific prediction and mark it as a marker for that prediction.

3.3 Results

In order to develop an explainable retinal screening model, we have trained a Convolutional Neural Network (CNN) model named "DenseNet-121" on 'APTOS 2019 Blindness Detection Dataset' with state-of-the-art accuracy and integrated the model in community telemedicine tool as described in Chapter 2. Then, we explored different techniques of achieving Explainability in CNN models and implemented a post-hoc technique named "Grad-CAM" on retinal images to achieve explainability.

3.3.1 Training a Convolutional Neural Network

To develop an explainable Retinal Screening model, we did some preliminary work by replicating the model used in Google Research [30] and have developed an Automated Assessment model using Convolutional Neural Network (CNN). As an architecture of CNN, we have selected Inception_V3, a 42-layer network used in our literature mentioned above. Although this model contains a much fewer number of parameters to train (24 million) compared to AlexNet (60 million) or VGGNet (180 million), it performs very similarly to VGGNet, with much more computing efficiency and a lower error rate. We have trained the model on the "APTOS 2019 Blindness Detection" dataset, which consisted of around 4500 images, each labeled with a number within 0-4 where 0 means 'no sign of DR' and 4 means 'severe DR'.

We have used Tensorflow 2.0 as the backend and have imported the Inception_v3 model from Keras, which was pre-trained on the "ImageNet" dataset. In pre-processing step, we have performed the "Sampling" and "Image Augmentation" techniques to balance the number of samples from each category. Due to constraints on computing resources, we executed only ten epochs, and with that, our model shows an accuracy of 53.01% on the test dataset. However, since this accuracy is not feasible, we have explored a few ways to improve this result. One of the options was to increase the dataset; another was to use a more sophisticated model. Thus, we used DenseNet-121 as our next model which promises to improve image classification. DenseNet-121 is a 121-layer network, designed to provide better accuracy with an even smaller dataset [35]. Instead of L connections in L layers, this network has L(L+1)/2 direct connections. The model's performance was measured by "sensitivity", which measures how many positive samples were correctly identified, and "specificity," which measures the proportion of correctly identified negative samples. With our trained model, we were able to achieve 90.6% sensitivity and 80% specificity on the test dataset.

3.3.2 Post-hoc Explainability using Grad-CAM

Though our primary goal is to make the screening model explainable to the graders, we cannot compromise the accuracy to gain that explainability. We plan to achieve state-of-the-art accuracy and include the explainability to the model. Thus, we have implemented a Post-hoc explainability technique named Grad-CAM [54] for the final output layer of the model and calculated heatmap for the input images. The generated Grad-CAM shows the heatmap around the DR retina symptoms, which can be very interpretable for the doctors and patients.

From Figure 3.2, it is showing the area which causes a retina having DR. Also, the image from the right is a Grad-CAM of a functional retina, which focuses on the vast central area not to find anything unusual relating to DR.

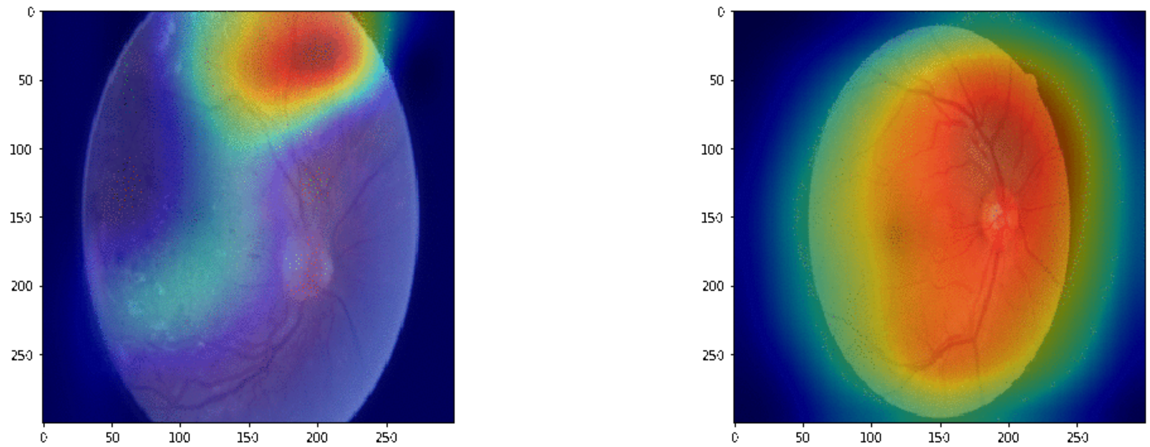


Figure 3.2: Grad-CAM for retina with and without DR

3.3.3 Evaluation Metrics for Explainability

The concept of explainability in machine learning models translates to our trust in the prediction, just as we trust a human grader. The underlying rationale is that explanations should be faithful to the mechanism of the underlying machine learning model. Interpretability/explainability has no strict or industry-standard definition; thus, we do not have any quantifiable metric to compare how explainable our model is. In the machine learning research area, there is an ongoing call for a degree of faithfulness. Hoffman et al. [34] has summarised key concepts of measurement of explainability and has identified some specific methods for evaluating:

1. the goodness of explanations

2. whether users are satisfied by explanations

3. how well users understand the AI systems

4. how curiosity motivates the search for explanations

5. whether the user's trust and reliance on the AI are appropriate

6. finally, how the Human-Explainable AI work system performs.

Zhou et al. [70] has conducted a survey on the quality of machine learning explanations and has concluded that the evaluation of ML explanations is a multidisciplinary research topic, and it is also not possible to define an implementation of evaluation metrics applicable to all explanation methods.

CHAPTER 4
Diabetes Self-management through Brief Motivational Interviewing

4.1 Managing diabetes through Motivational Interviewing

This section covers a brief discussion about the definition of Motivational Interviewing (MI), the 'spirit' of MI, brief MI (a particular form of MI), and finally about Brief Action Planning based on the work published by Gutnick et al. [31]. It also includes a discussion how an electronic tool based on the concept of 'Brief Action Planning' can reduce training time for clinicians and can facilitate action planning process, thus, supporting the self-management process of diabetic people.

4.1.1 Motivational interviewing (MI)

Motivational Interviewing (MI) is defined as a "collaborative, patient-centered form of guiding to elicit and strengthen motivation for change" [42]. Practitioners of MI adhere to the Spirit of MI (compassion, acceptance, partnership, and evocation). "Compassion" indicates a primary concern for the patient, as opposed to one's self-interest. "Acceptance" refers to empathetic communication, "Partnership" refers to a sense of equality and reciprocal relationships in the encounter, and "Evocation" suggests that ideas for change are most effectively elicited from the patient, rather than given to the patient by the clinician.

4.1.2 Brief MI

There is no authoritative or widely accepted definition of "Brief MI." PubMed searches reveal over a dozen different forms of "Brief MI," developed, tested, or used in a wide variety of settings. For this article, "Brief MI" will be used to refer to any standardized formal set of interventions intended to achieve similar behavioral outcomes as does traditional, standard MI.

4.1.3 Brief Action Planning

Brief Action Planning (BAP), a form of brief MI, is defined as a highly structured, stepped-care, evidence-informed, self-management support tool and technique based on the principles and practice of MI [31, 51].

4.1.4 Components of BAP

BAP consists of eight core competencies, often referred to as three core questions and five skills, all used with the spirit of MI. Reims et al. [51] have discussed the rationale and evidence supporting the questions and skills used in BAP. A summary of all the components has been included in this paper to help readers better understand BAP's fundamental approach and how it helps patients develop achievable action plans.

Question 1: Both focuses and evokes action planning To bring the focus of the conversation to the development of an action plan for self-management of behavioral

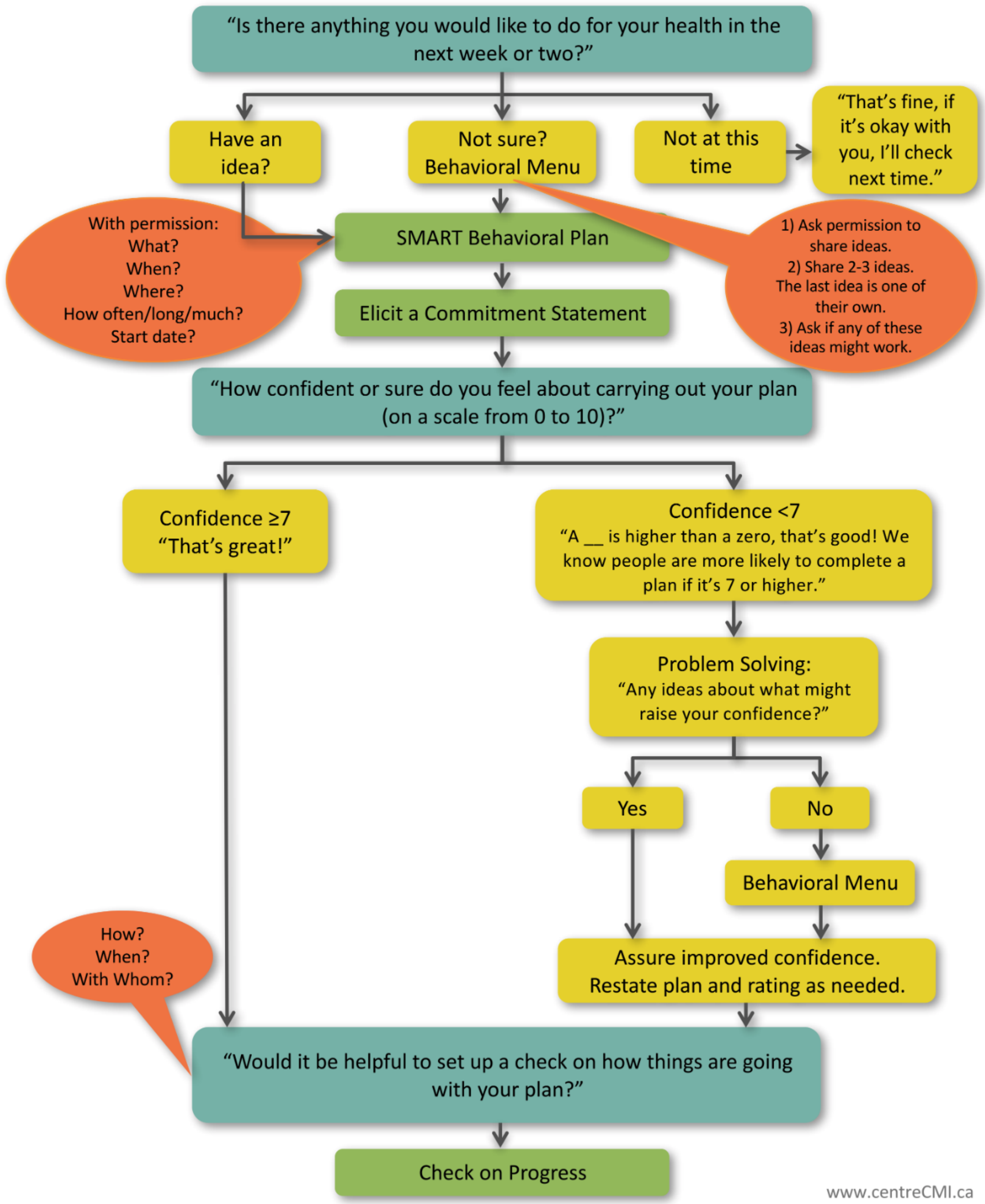


Figure 4.1: Flowchart showing the components of Brief Action Planning (citing from Gutnick et. al. 2014)

change, the very first question of BAP asks, *"Is there anything you would like to do for your health in the next week or two?"*. Clinicians ask this question once rapport has been established, and they think that the patient may be ready for self-management planning. There are three types of responses to this question, and depending on which direction the patient chooses, the conversation takes different paths.

- **"Yes or I have an idea:"** When the participant immediately has an idea about what they want to do, clinicians proceed to the following competency: SMART planning.
- **"Not sure:"** When the participant wants or needs suggestions, clinicians can utilize a behavioral menu (another core competency) after asking and obtaining permission from the patient.
- **"No or Not at this time:"** A patient may be unwilling or unable to make an action plan. Consistent with the Spirit of MI, which respects participant's autonomy (to change or not to change), clinicians can then encourage patients to return to a discussion about change whenever they might become interested.

Skill 1: Offering a behavioral menu

A behavioral menu is offered to a patient by the clinician when the patient shows a desire to discuss action planning but seems unable to formulate his/her ideas about an action to pursue. Three distinct evidence-based steps are followed while offering the behavioral menu:

- Ask permission to offer a behavioral menu
- Offer two or three general yet different ideas, all at once, from which a patient may choose one for himself, or be inspired to develop an idea of his/her own. It is crucial to maintain the language to imply that the suggestions trigger ideas from the participants, not to give the perfect solutions [42]. One example can be *"One patient I work with decided to join a gym and start exercising, another decided to pick up an old hobby he used to enjoy doing, and another patient decided to schedule some time with a friend she hadn't seen in a while."* [31]
- Ask the patient whether any of the presented ideas might work for them or if the participant has come up with a new of his/her own in the meantime.

Skill 2: SMART Planning

On receiving an affirmative response from Question 1 or after offering a behavioral menu that results in, "something they would like to do", the facilitator works with the patient to develop a SMART (Specific, Measurable, Achievable, Relevant, and Time-bound) plan. Literature [11, 40] has shown that people are more likely to complete an action plan when the plan is specific, proximal, and achievable rather than being vague and over-ambitious. The SMART plan can be accomplished through a series of questions relevant to the plan, such as, *“what?” “where?” “when?” “how long?” “how often?” “how much?”* and *“when do you want to start?”*

Skill 3: Elicit a Commitment Statement

After developing the specific plan, the facilitator asks the participant to repeat the entire plan altogether by saying something similar to *”Just to make sure we understand each other, would you repeat back what you have decided to do?”*. This act of repeating back helps the participants to do a thorough review of the plan in their mind and perform a quick feasibility analysis. A person’s sense of feasibility is often reflected through the strength of commitment language used at this step [53, 7]. If accurate, the clinician affirms that is what they also recorded in the plan’s development and might say, “Great, that is also what I heard.”. If incomplete or inaccurate, there is an opportunity for correction.

Question 2: Scaling for Confidence

After eliciting the commitment statement, the facilitator assesses the participant’s self-efficacy to complete the plan through the next question *“I wonder how confident you feel about carrying out your plan on a scale of 0-10, where 0 refers to “no confidence at all” and 10 refers to “total or full confidence”. ?”* [42].

Skill 4: Problem Solving for Low Confidence

If the patient’s confidence is less than 7 out of 10, a collaborative problem-solving approach is suggested as scores less than 7 out of 10 are associated with lower likelihoods of completing plans [40]. The first step of this collaborative approach is to affirm that any confidence level is better than no confidence at all. The core concept is to emphasize a "strength-based" approach vs. a "deficit-based" approach, with three core elements:

- Emphasize the positive, not negative

  (For example, "A level of confidence of five indicates a strong interest in changing. After all, you could have said a three or even a one.")

- Provide background and evidence supporting an effort to seek higher number

  ("We know from experience and research that a confidence level of seven or greater is associated with a greater chance of being able to complete your action plan."

- Invite collaborative problem-solving

  ("Do you have ideas of how you may be able to increase your confidence in your plan?")

The facilitator can offer a behavioral menu at this step. With consent from the participant, the facilitator can share two to three ideas that have helped the participants improve their confidence level.

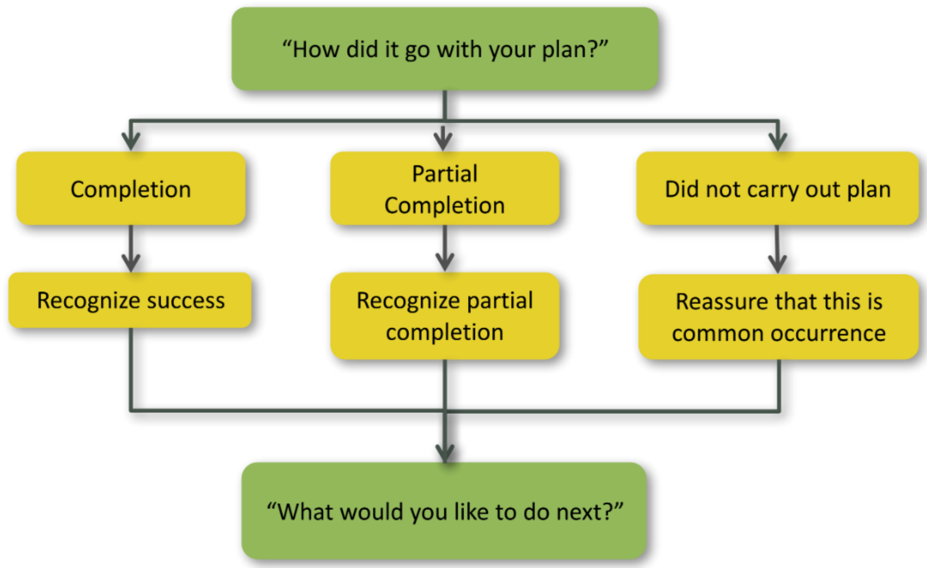


Figure 4.2: Flowchart of follow up meeting (citing from Gutnick et al. 2014)

Question 3: Arranging Accountability

After a confidence level is addressed, the next question is *"Would it be helpful to set up a check on how things are going with your plan?"* which facilitates the patient to arrange accountability for their plan. People are more likely to complete a plan when they feel accountable and include a follow-up or accountability for their progress[17, 50]. Accountability could be with anyone, such as the clinician himself/herself or any supportive person(e.g., family, friends, etc.) or even themselves (e.g., set a calendar invite to check in with the plan).

Skill 5: Check on Progress

Follow-up is one of the prominent features of successful self-management action plans[40]. The flowchart shown in Figure 4.2 depicts the possible outcomes measured during a follow-up session. The facilitator asks the patient about the plan and how it went, and whether they are interested in making any changes or a new plan altogether. This next step can be to continue the same plan, make minor/major adjustments to the plan under review, or even create a new plan. If the patient decides not to have next step, the facilitator prioritizes the patient's autonomy and moves on to the next part of the visit.

4.1.5 How an electronic version could reduce the training time of MI and make it more available to physicians

While using BAP in clinical settings, paper forms' clerical burdens have made it non-editable, error-prone, and difficult to analyze, which is crucial for the future improvement of the BAP model. The benefit of digitizing the paper forms has already been established through literature [43], specifically after signing the Health Information Technology for Economic and Clinical Health (HITECH) Act into law 2009. Patient care can be improved significantly when Electronic Health Records (EHR) are designed and adopted in a "meaningful" way [41]. However, for a patient-centered communication environment like BAP, a traditional EHR is not adequate. Capturing the interaction between patients and clinicians during an interview is challenging since no single approach or flow of conversation works for everyone. Moreover, to utilize the techniques of BAP in regular healthcare systems, we need to develop an electronic version of the tool that can guide the clinicians to make an effective action plan with minimal training.

4.2 Overview of CAP System

Following these challenges of using BAP in clinical settings, in this chapter of the dissertation, we address the subsequent research questions:

RQ1: How can we design a framework to support patient self-management and maintain the fidelity of BAP?

RQ2: Is CAP an acceptable clinical intervention?

RQ2: Is CAP able to reduce training time for the facilitators while maintaining fidelity of BAP?

We propose a unique web-based software framework, referred to as CAP, which can facilitate and enhance the implementation of BAP in clinical settings. It can be extended and scaled for any other intervention used by the clinicians to encourage behavior change. The framework has been designed to help clinicians maintain the fidelity of BAP and adjust all possible directions a conversation can take during an interview session for behavior change. On top of all basic features required in a traditional EHR, it includes several facilitating features. Among all, prompt to the clinician depending on the stage of conversation using MI conversational style helps maintain "fidelity of the intervention". It supports clinicians working to utilize these skills for in behavior change management, potentially reducing the cost of training. The CAP design is a collaboration of

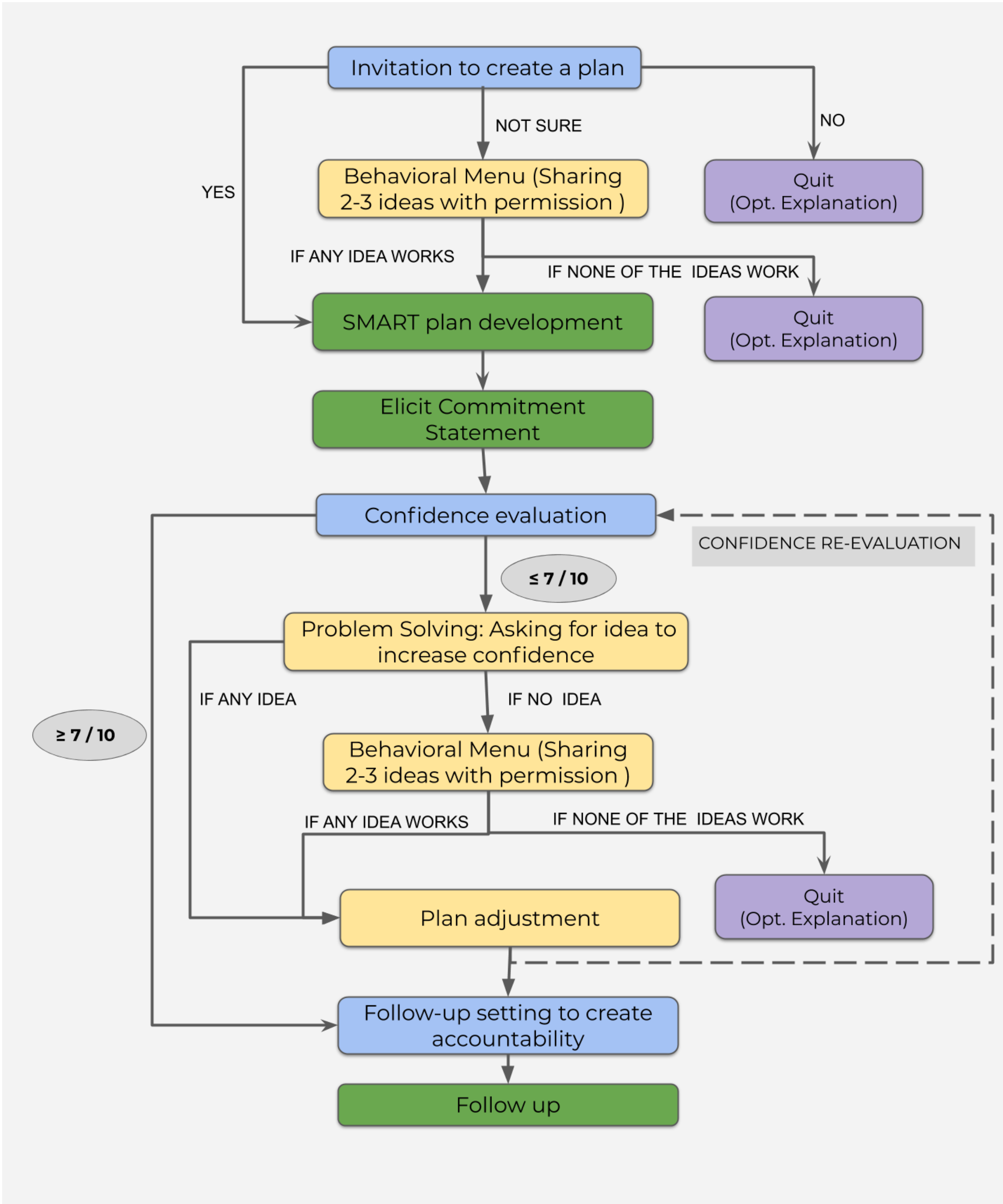


Figure 4.3: Questionnaire Flowchart in CAP following Brief Action Planning

Computer Science with Psychology and Behavioral Science to contribute towards the emerging area of Connected Health.

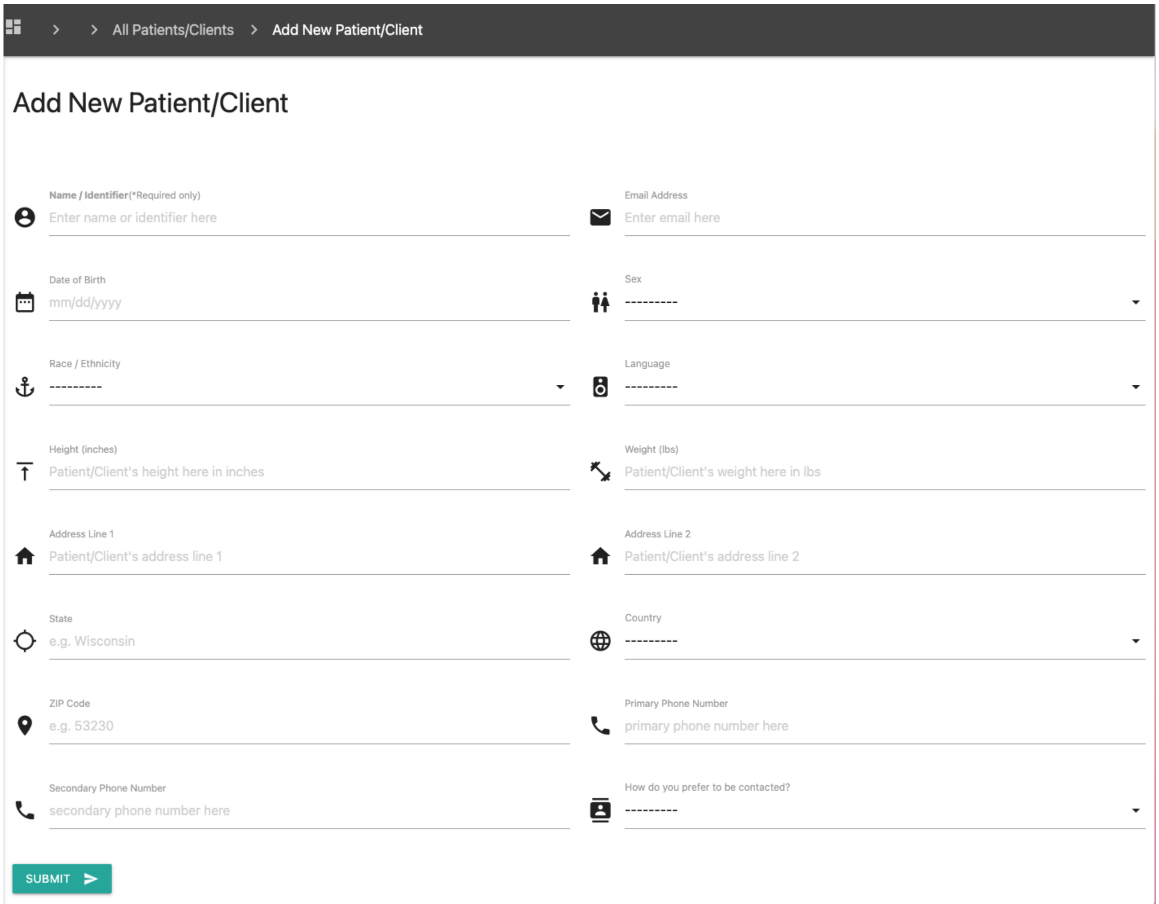


Figure 4.4: Screenshot of 'Basic Feature: Add new participant' from CAP system

4.2.1 Basic Features

Add new participant

Facilitators can add a new participant with this feature. They need to insert information from a participant and submit as shown in Figure 4.4. Only the first name of the participant is required to add a new participant.

View All participants

Facilitators can view all participants added by them using this feature. They can view plans and profile of a participant and also edit a participant's profile.

Edit participant

Facilitators can edit a participant's profile from this page. Previously inserted information can be edited as well as any unfilled information can be added.

View profile

Facilitators may need to see an individual participant's profile for contact info or address. These can be viewed on the 'View profile' page.

View All plans

All plans of a participant can be viewed on this page. Facilitators can see add, view, edit or follow-up plans from this page.

Add new plan

Facilitators can add a new plan for a participant. Currently, this feature supports two languages: English and Spanish. The questions follow the flow of Figure 4.3. Depending on the responses to the very first question, the following questions will vary. As described in the section covering the questions and skills used in 'Brief Action Planning', three different scenarios might occur. If the response to the question "Is there anything you would like to do for your health in the next week or two?", if the response is 'yes or possibly yes', a SMART plan is developed through detailed and specific questions about the plan as shown in Figure 4.5a. If the response is 'Not sure', the behavioral menu is utilized which can be followed by SMART plan development, if successful, as shown in Figure 4.5b. Finally, if the response is ‘No’, the autonomy of the patient is respected and the plan is submitted as shown is Figure 4.5c

View a particular Plan

A particular plan of a participant can be viewed on this page. Facilitators can edit the plan from here. They can also email and print the plan details.

Edit Plan

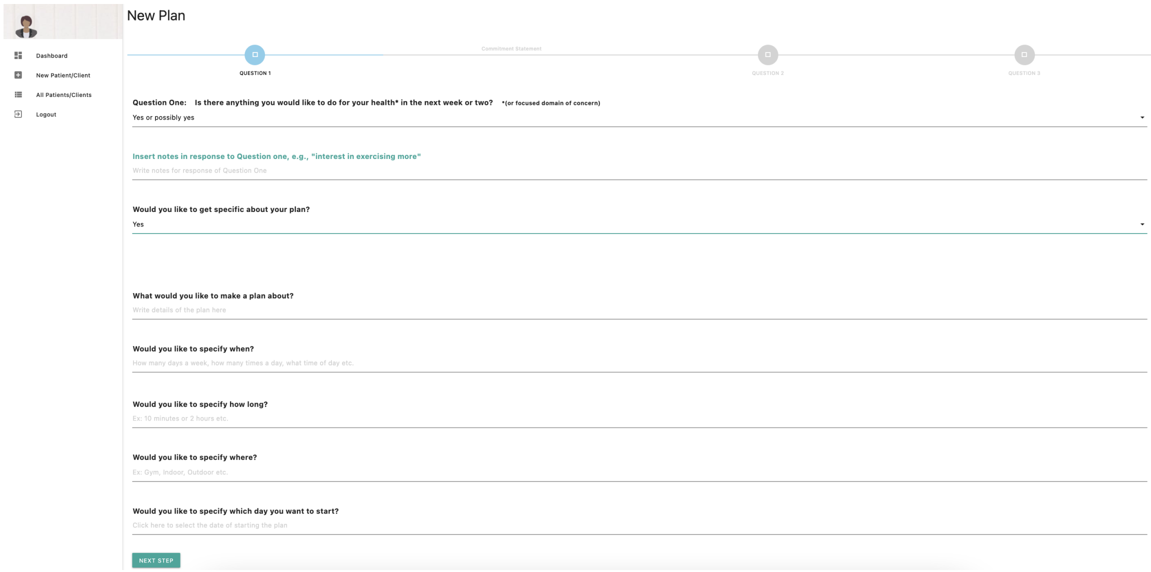
Plans of a participant may be changed. Facilitators can edit them using this feature.

Create a Follow-up

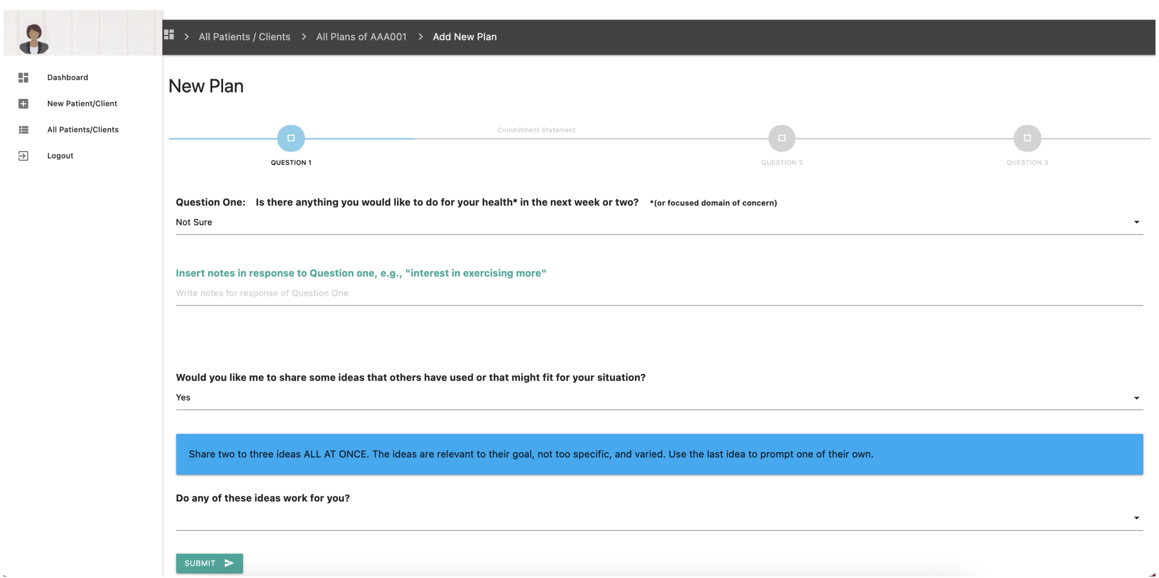
Facilitators can check on the progress by following up with the participants. Participants can set up the next follow up time and date when they add or edit a plan. Figure 4.7 shows how follow-up is designed in the CAP framework. Figure 4.6 shows the flow chart of the Follow-up questionnaire. The facilitator asks the patient about the progress of the plan. The progress of the plan is measured by the following criteria.

- Completion - 50% or more of the action plan is achieved.
- Partial Completion - 1-50% of the action plan is achieved
- Did not carry out the plan - Otherwise

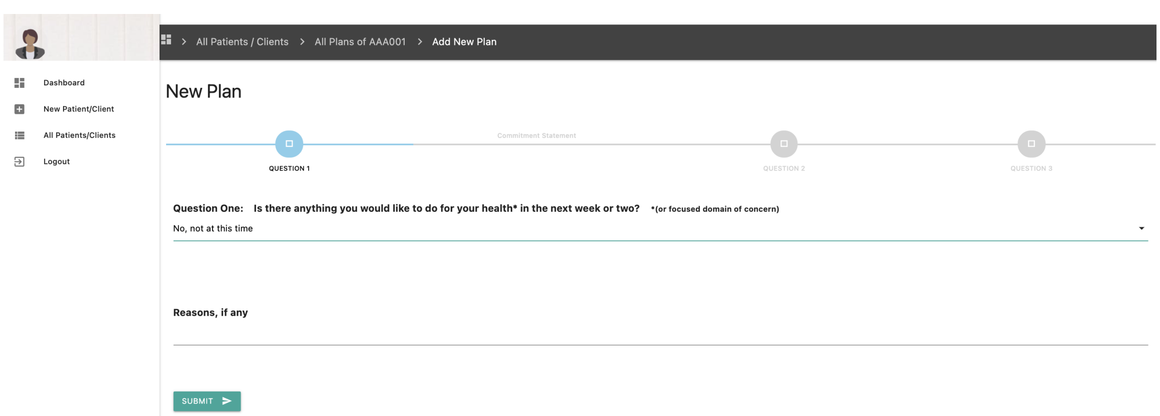
Prompts are being shown upon the selection of the follow-up options. These



(a) Scenario 'Yes or possibly yes' of 'Basic Feature: Add new plan'



(b) Scenario 'Not sure' of 'Basic Feature: Add new plan'



(c) Scenario 'No' of 'Basic Feature: Add new plan'

Figure 4.5: Screenshot of 'Basic Feature: Add new plan' from CAP system

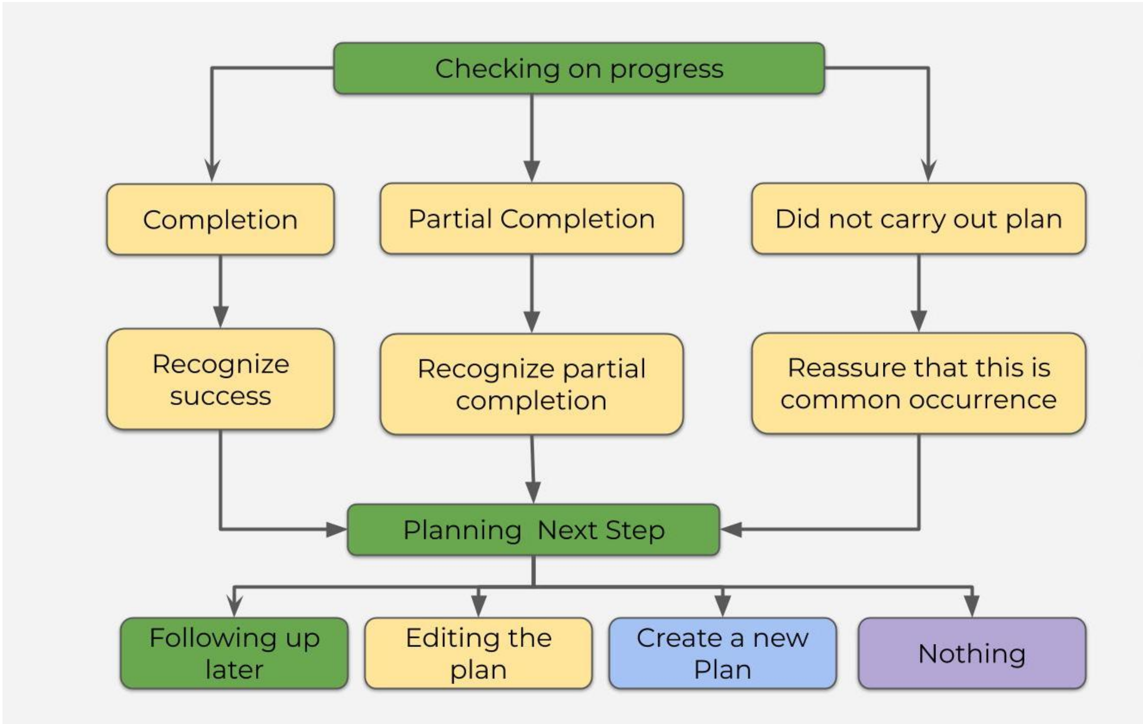


Figure 4.6: Questionnaire Flowchart of Follow up meeting

prompts help the facilitators to keep focused on the participants and also motivates the participants. If the plan is completed, then the prompt is *"Good for you! Completing your whole plan shows a strong commitment to your health."* If the participant partially completes the plan, this prompt can motivate the participants - *"Good for you! It's great that you were able to complete some of your plan, which is more than most people."*. If the participants did not carry out the plan, the following prompt is shown - *"That's OK, it is normal for people to make goals that are difficult to achieve and that might not work out"*.

After acknowledging the response from participants, the facilitator will ask the participants about how they would like to proceed. The participants can choose

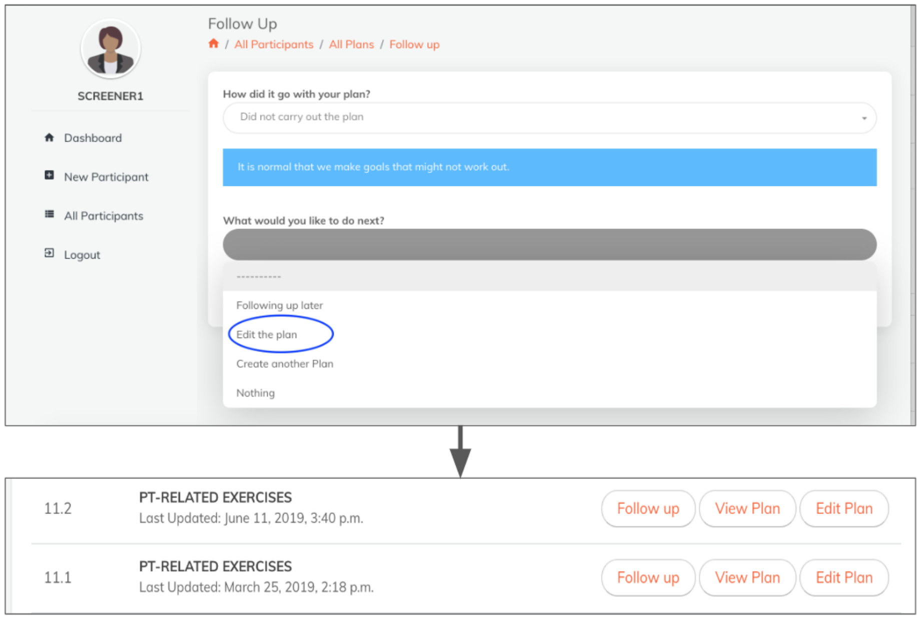


Figure 4.7: Screenshot of CAP system showing the plan versioning feature

to follow up later, edit the current plan, create a new plan, or do nothing at all. Facilitators respect the consent of the participants and respond accordingly.

4.2.2 Facilitating Features

Versioning of Plans:

One of the most thoughtfully designed components of CAP is the support of multiple versions of the plan.In the follow-up meeting, after checking with the progress, the facilitator asks the participants what they want to do next. While they are offered the options like "following up later", "edit the previous plan", "creating a new plan" or "Nothing". If the participants want to edit their previous plan, then the edited plan is saved as a new version of that plan as shown in Figure 4.7. This feature supports the opportunity to perform analysis of data gathered

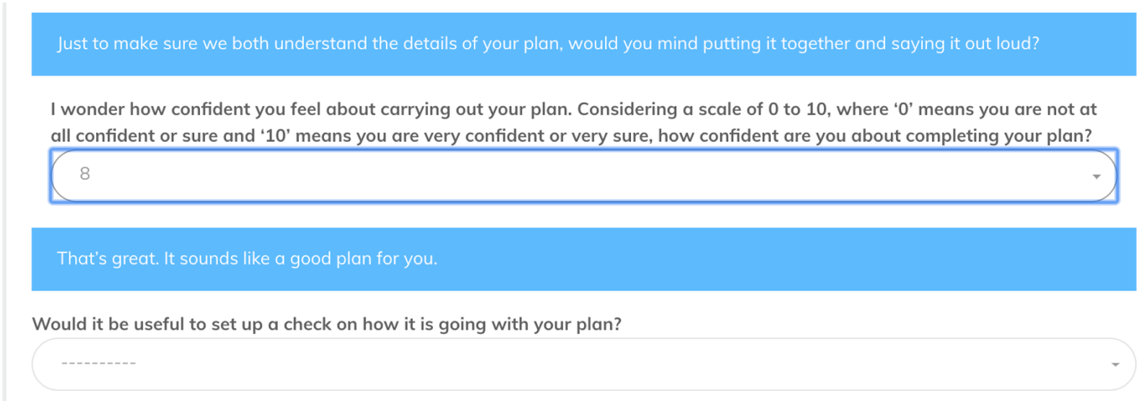


Figure 4.8: An example of prompts guiding CAP interview

from different versions and completion rates to identify the factors that helped the participants achieve their goals.

Useful prompts depending on response

As shown in Figure 4.8, CAP provides some prompts depending on the direction of the conversation aiming to facilitate the interview process by helping the facilitator to adhere to the language emphasizing the spirit of MI. It also helps to keep the focus on SMART goals and other components of BAP. O'Leary et al., in their work [47] about Designing Peer Support Chats for Mental Health, have shown that chats guided through prompts were more helpful and more directed to solutions.

Free text-box to capture explanations related to individual responses

Out of respect the patient's autonomy, CAP allows the participant to quit the planning process at any point during the interview. In such scenarios, when the facilitator directs the conversation towards an end, an optional free-text box is

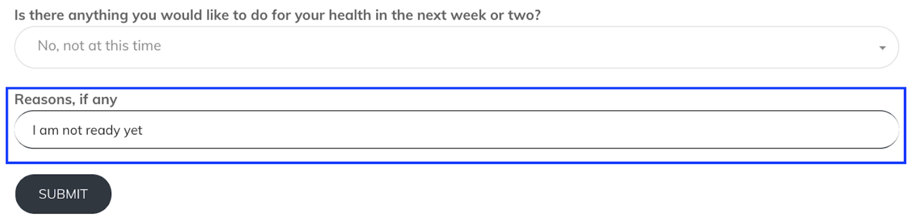


Figure 4.9: Screenshot of CAP system showing the free text-boxes

presented in CAP to capture the reason, if any, expressed by the participant as a reason for stopping the process.

Printing and Emailing a copy of plans to the participant

This feature is included in CAP according to the design considerations collected from real-life clinical scenarios. As the plan is captured and recorded in CAP, providing the printed copy to the participants would allow them to review and follow through the plan. The email feature improves this scope and helps to have a more accessible electronic version. Moreover, upon plan completion, if a facilitator option is selected for follow-up, an automated calendar invitation is sent to the facilitator's email to ease calendar integration and support patient management.

4.3 Study on Acceptance of CAP among clinicians

After completing the development of the beta version of the CAP framework with the features described in the previous section, we have conducted a study to evaluate acceptance of CAP as an clinical intervention tool. This section covers design and motivation of this study followed by the results and discussions.

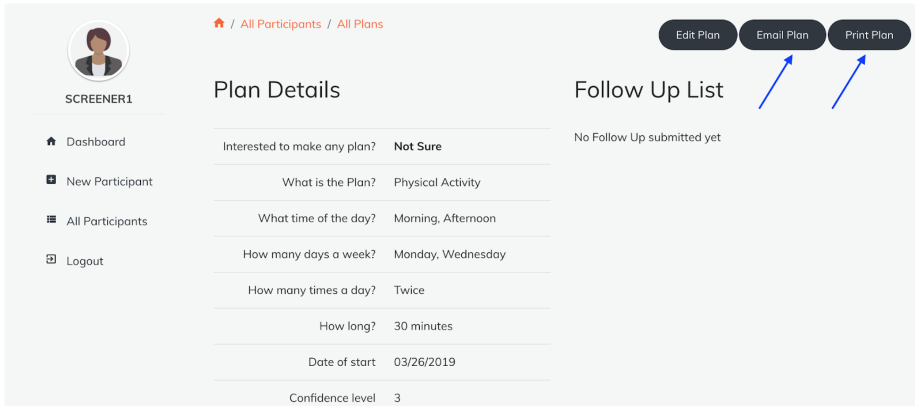


Figure 4.10: Screenshot of CAP system showing the Print and Email feature

4.3.1 Study design

We designed an evaluation study to understand acceptance of CAP among clinicians who facilitate behavior change. For the study design, we have combined two methodologies from literature [39], such as Walkthrough demonstrations and Likert Scale Questionnaires, to elicit user feedback. Our goal was to introduce the CAP framework to the clinicians and obtain feedback on its design, whether they would use it in their routine clinical care and to gather feedback on improvements of the CAP framework for effective integration into existing EHR systems.

The questionnaire consists of ten questions in three sections covering feedback on the overall efficacy of the software, then about the intuitiveness of the software design and finally feedback on the usefulness of some specific features. At the very beginning, few non-identifiable demographic questions like profession and age group were included for the sake of the study. The participant's baseline technical proficiency is asked to determine if there is any possible bias in the responses about design intuition. The first section is about “Facilitation”, i.e., it was designed to ask questions on whether and how CAP facilitates the clinicians to conduct an effective interview using BAP. The next section, named “Design Intuition” contains questions about the intuitiveness of the design of CAP. And, finally, the “Specific feature usability” section includes questions about the efficacy of the additional unique features that have been incorporated in CAP by collecting requirements from the direct experiences of health professionals in the clinic.

4.3.2 Data collection

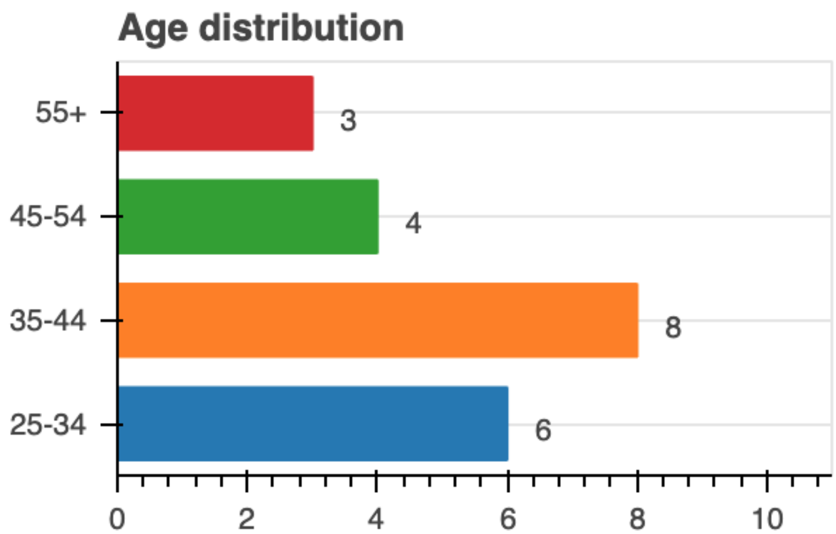
We have performed a walkthrough demonstration among eighteen physical therapists at Milwaukee VA Medical Center in Milwaukee, Wisconsin, and then conducted a follow-up survey to understand what they thought about the CAP framework. Similarly, we directed a similar session at the Medical College of Wisconsin in Milwaukee, Wisconsin, with four physicians, two formally trained in MI and two untrained in MI, who develop action plans with their patients. During the one hour session, we had a short demonstration on the idea of Brief Action Action followed by a walkthrough of CAP, and it’s significant features. The clinicians showed a very positive response and interest in BAP and volunteered to use the framework by arranging a demo session where one clinician performed the role of a patient, and another clinician acted as a facilitator. The facilitator used the CAP framework to create an action plan with the other person performing as a patient. As in BAP training, researchers encourage 'real play' where possible, the participating clinicians worked on real action plans for their own health. Our immediate observation was that they needed very little to no help using CAP even though they were using it for the first time. One group was able to switch their role after completing one full session, i.e., they exchanged the patient and clinician's role and made another action plan. At the end of the session, we conducted the follow-up survey, the same one used for Physical Therapists in Veteran Affairs and we had three completed surveys from this session.

4.3.3 Results

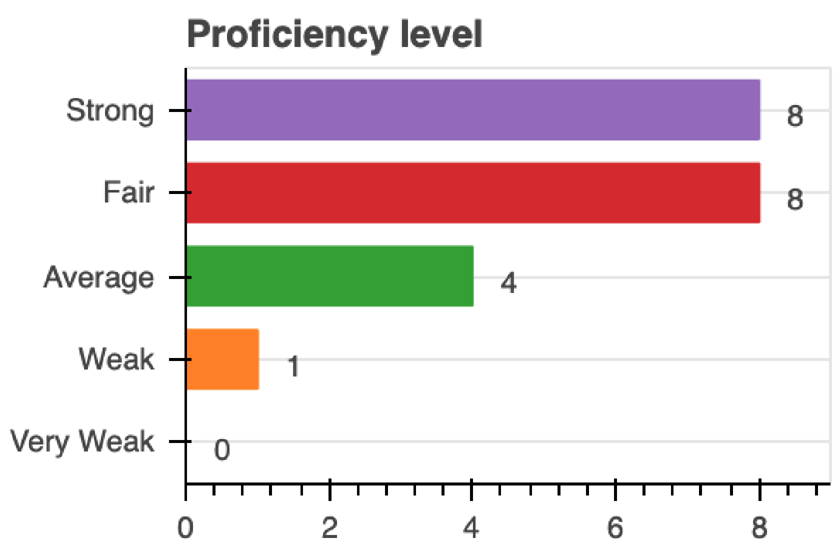
We have analyzed the responses from 21 clinicians (18 physical therapists and three physicians) who, at least once in their professional life, have directly worked with their patients to develop a health goal. We have two questionnaires with partial responses. We have excluded only the unanswered questions for those. For clarity, we presented the results under different categories in consistency with the sections of the questionnaire.

Demographic Overview

From the age distribution plot in figure 4.11a, we can infer that the clinicians were well-distributed in terms of age, and there was no age-related bias in the responses. As a reference point for our evaluation regarding the intuitiveness of the CAP tool,



(a)



(b)

Figure 4.11: Age distribution and Proficiency level of clinicians in using technical platforms

we have asked the clinicians to rate their technical proficiency in using any new software. For unbiased interpretation, the categories were explained in the questionnaire with example cases:

Strong : I figure out all functionalities with little effort all by myself

Fair : It takes effort, but I get all the features by myself

Average : I can use basic features, but often need training for advanced features

Weak : I need training for basic features, but am competent with basic features once trained

Very weak : I struggle with basic features even after training

The distribution of technical proficiency of the clinicians has been plotted in figure 4.11b, and only one out of 21 has marked them as weak, whereas eight marked them as Strong, eight as Fair and four as Average.

We have also conducted a one-way between-subjects ANOVA to compare the effect of technical proficiency on the rating of intuitiveness of the tool. Our analysis did not find any statistically significant difference between groups with varying technical proficiency.

Results: Facilitation

The clinicians were asked, "How helpful was the CAP software to facilitate the action plan process?" and the responses were mostly very positive, as shown in the

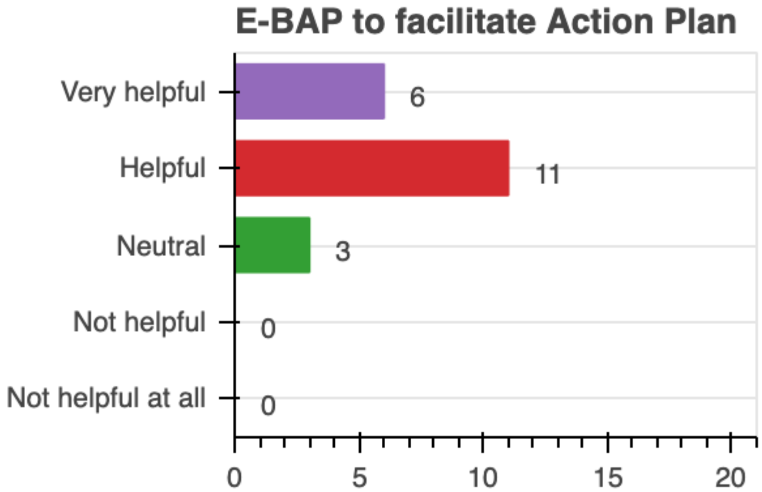


Figure 4.12: Responses for question about facilitation of CAP

graph 4.12. When they were asked to rate the usability of CAP on a scale from 0 to 10 (where is 0 is lowest and ten is highest), we have found the mean score to be 7.6.

Results: Design Intuition

This part of the questionnaire focused on what the clinicians thought about the design intuition of the CAP. The Likert chart had the following categories with clear explanations:

Very intuitive: I could figure out all functionalities very easily without any help

Somewhat intuitive: It took some of my effort to figure out the available functionalities

Average: I could figure out most of the features and ask for a few features

Not that intuitive: I had to ask for help for most of the features

Not intuitive at all: The design did not make sense to me)

Upon analysis, 55% of the clinicians found the design to be 'very intuitive,' 40% marked them as 'somewhat intuitive,' and 5% were neutral on this question. For the rating of intuitiveness in a 0 to 10 scale (0 is lowest, and ten is highest), the mean score was 7.94, with 8 being the highest frequent rating number.

Results: Feedback on specific features

There are some particular features integrated into the CAP which has been designed to facilitate the interviewing process. The motivation behind these features' design requirements has been discussed in the "Design and Methodology" section already. However, in the questionnaire, the last part was about gathering feedback on clinicians about these specific features.

About the prompts shown during the interview according to the flow of the conversation, the clinicians were asked *"CAP provides some useful hints (in blue colors) to the facilitator depending on the response of the participants. How helpful do you find these prompts during an interview?"* along with a screenshot of this specific feature in CAP. It was a great pleasure to find that 100% of the clinicians have marked them as "Very helpful," which is shown in Figure 4.13.

As a follow-up question to this response, they were asked *"If you selected the* *prompts(in blue color) in previous questions to be helpful, why did you find it helpful?"* and this question had the following categories:

- It speeds up the goal planning process
- It helps to keep the focus on SMART goals and evidence-based behavioral components.
- It facilitates the use of language to emphasize the patient's autonomy, empathy, etc.
- I did not find it to be helpful
- A free text-box to capture any other response.

We have seen that 50% clinicians have responded with the third option, i.e., it facilitated their use of language emphasizing patient's autonomy, empathy, etc. and 35% selected the second one, i.e., it helped them to keep the focus on SMART goals. 15% of them reported that the prompts had facilitated them to speed up the action planning process.

Regarding the inclusion of optional text-boxes in CAP, the clinicians were asked *"CAP provides few optional free text-boxes for capturing an explanation that the participants provide anytime they want to quit making the plan. How helpful can these texts be to find out the motivating/demotivating factors of goal-setting for*

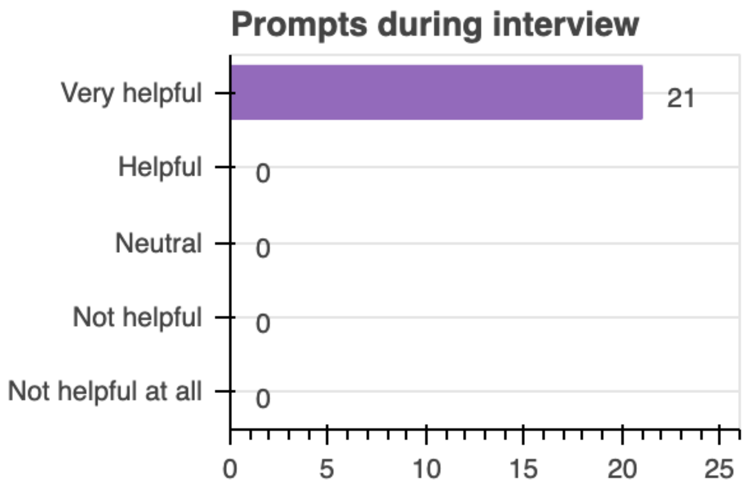


Figure 4.13: Feedback on the prompts shown during interview

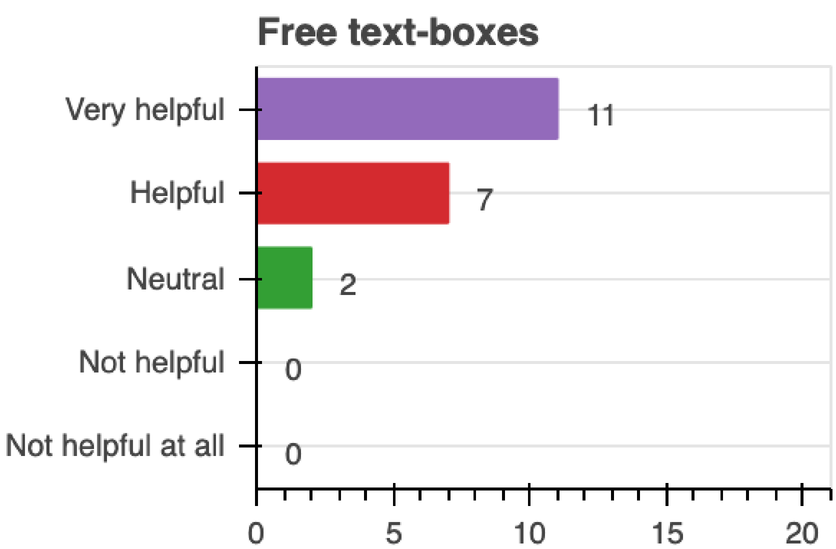


Figure 4.14: Feedback on the free text-boxes on all exit points

participants?". The response was entirely distributed compared to the responses about prompts. 55% clinicians marked it as "Very helpful" while 35% marked it as "Helpful" and 10% were neutral about it as shown in the plot 4.14.

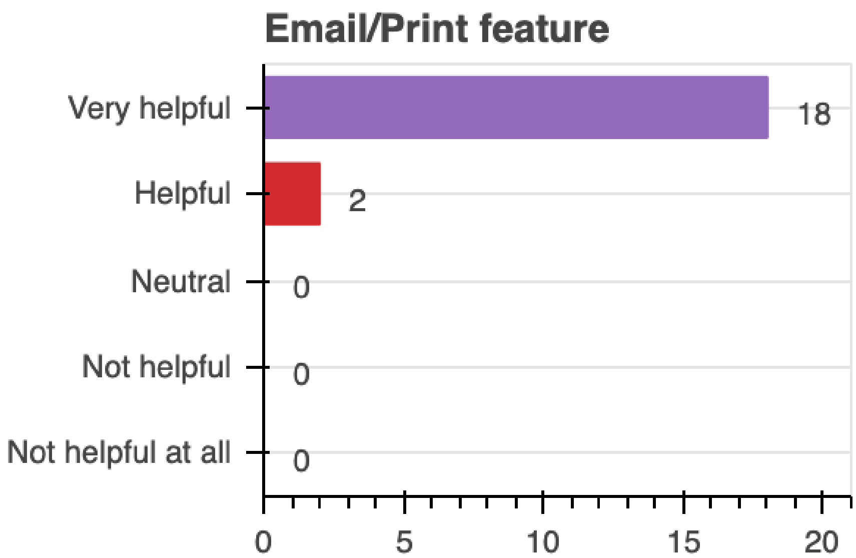


Figure 4.15: Feedback on the Emailing and Printing feature

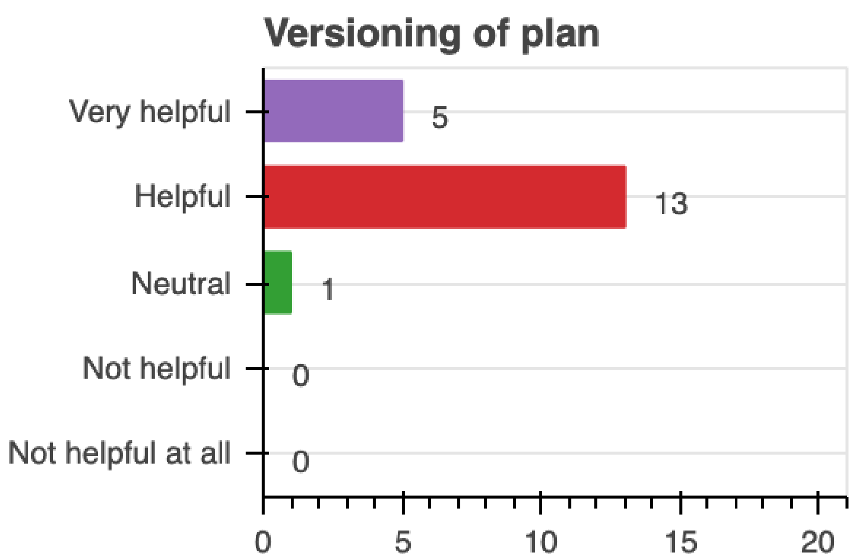


Figure 4.16: Feedback on the version support of Action Plans

About the Emailing and Printing feature of CAP, 90% clinicians have responded with "Very helpful," and 10% responded with "Helpful," which also summarises this feature to be positively accepted as shown in Figure 4.15.

In the follow-up meetings, CAP allows editing the previously built action plans and saves that as a new version of that plan. As shown in Figure 4.16, only one clinician was neutral about this while the rest of them have marked this feature to be helpful and very helpful.

4.3.4 Discussion

Based on the acceptance evaluation, the results show that it has been well-accepted by the clinicians who took part in the survey. The clinicians also had the opportunity to express their thoughts as a response to some open-ended questions like whether they had any particular feedback on the CAP tool, e.g., what feature needs to be added/changed/removed or any other impression, how it can be made more intuitive and user-friendly, etc.

A few responses included some suggestions related to the User interface. For example, one clinician thought that enlarging the font used in the interface could help the clinicians. Another clinician expressed that the prompts would look better if they were bold and larger font. A few suggestions have already been included in the following iteration of CAP. One of the major concerns that we found among the clinicians is that those who are already using any Electronic Health Record (EHR) for the data management of their patients are not willing to use any standalone new platform for BAP. Instead, they recommended including CAP into the commonly used EHRs like Epic. CAP is designed in a modular pattern at the architecture level, and it can be integrated into any other platform by using the Application Program Interface (API) of CAP. The final thoughts of the clinicians indicate that CAP can be adapted in any sector of Health care, particularly the sectors related to chronic diseases, whenever it requires the patient to work on self-management by developing an action plan.

4.4 Study on Efficacy of CAP in reducing facilitator training time

We designed and conducted another study to determine if CAP is able to reduce training time for the facilitators while maintaining the fidelity of the intervention. This section covers a brief description of study design, data collection, data analysis, and finally, the results and discussions from the study.

4.4.1 Study Design

The study was conducted among 40 students from the Department of Physical Therapy at Marquette University who were enrolled in the PHTH 4512 course during Fall 2020. The students had only a brief introduction to Motivational Interviewing. The study was exempted from Institutional Review Board at Marquette University as it fell under the 'Education Research' category. Moreover, no identifiable information about any student was part of this study. The goal of the study was to evaluate if students were able to maintain the fidelity of intervention with minimal training. With that goal in mind, a "Skills Checklist" was adapted from the Center for Collaboration, Motivation and Innovation containing a total of 18 questions as shown in Figure 4.17. Among those 18 questions, 16 questions covered the core components of BAP. Two questions covered if the trainee facilitators could maintain a warm and encouraging tone during the intervention and maintain the structure of the intervention. Each questions had 3 possible answers: a) 'Achieved', b) 'Developing' and c) 'Not applicable'. Depending on how many questions got marked as 'Achieved,' a total fidelity score was achieved by calculating a percentage of 'Achieved' with total applicable questions in the checklist. If more than two elements out of 16 questions covering the component of BAP were marked as 'Developing' for any student, it was considered that the facilitator (the student in this case) needs retraining and is not ready to conduct intervention yet. Moreover, if the questions covering the 'warmth and tone' or 'structure' were marked as 'Developing,' it was also considered a case for retraining.

4.4.2 Data Collection and Analysis

As part of the PHTH 4512 course, the students were given a training of one hour on principles of BAP and on how to use our CAP tool. Then, they were given a voluntary and ungraded assignment where every two students made a pair, and in each pair, one student used the tool as a facilitator to make an action plan for another student. They recorded this session and submitted the audio recordings of their session as an assignment.

Two independent reviewers who were trained on fidelity assessment of

CAPPA Skills Checklist

Date of Review: ____________________ **Participant ID:** ____________________
Reviewer: ________________________ **Study Staff Member:** ______________

Item	Description	A	D	NA
Question 1	"Is there anything you would like to do to increase your activity in the next week or two?" is asked clearly and respects the person.			
Skill 1: Behavioural Menu	The coach **asked permission** to offer a Behavioural menu if the participant doesn't have any ideas, doesn't know where to start, or requests ideas.			
	The coach offers two or three brief (not too specific) ideas to increase physical activity together in a list without pauses. The list has variety.			
	The coach asked the person if they had any ideas of their own as the last item on the list.			
Skill 2: SMART plan	The coach asks permission to make a detailed plan			
	The coach completed relevant SMART planning elements (What, When, Where, How often, How much, How long, Start date), if participant is willing			
Skill 3: Commitment Statement	The coach asked the client to say back their plan and indicates on *www.myactionplans.com* site that all elements in SMART plan were clearly articulated.			
	The coach gently corrects any SMART element missed during commitment statement and asks again for the plan to be repeated, "I also heard you mention [*that you specifically planned to walk at lunch*] in your plan was . . . Can you repeat the entire plan again just so we're both clear? I'd like to hear it in your words, if possible."			
Question 2	The coach asked confidence (how sure) level clearly with a description of what confidence and the numbers mean or provided a culturally appropriate alternative.			
	The coach responded positively to the person's confidence level and if the confidence level was below 7, explained the reason to revise plan for a confidence level of 7 or above.			
Skill 4: Problem solving	The coach used problem solving if confidence was less than 7.			
	The coach asked for the person's own ideas first.			
	A behavioural menu was offered (as above) if participant doesn't have ideas.			
	The coach asked for the commitment statement and confidence level again after the plan if the plan was altered. The updated commitment statement is not required, but recommended.			
Question 3	The coach asked for permission to include an accountability plan.			
	The accountability plan was clear, specific and determined by the person. (with whom, how, when)			
Warmth and tone[1]	The tone is warm and encouraging, and the participant does most of the talking. There may be statements of encouragement such as "that sounds like a plan that will work for you," and the coach does not use language or statements that reinforce an 'expert' role.			
Structure	The items occurred in the order that they appear on the checklist.			
Total fidelity score = ___ / 18 elements. If greater than 2 elements missing schedule for retraining If Warmth and tone or structure are developing schedule for retraining.				

***A**=Achieved; **D**=Developing; **NA**=Not Applicable [1]The Spirit of MI (compassion, acceptance, partnership, and evocation) is built into BAP.

Figure 4.17: The skill checklist used to calculate fidelity score

Motivational Interviewing graded those audio recordings of the student's sessions. For the sake of reliability, both reviewers initially graded five sample recordings, and then Cohen-Kappa Score [62] was calculated based on those five sample grading. Finally, 40 recordings submitted by the students were graded, and the results from the analysis of that data will be discussed in the following subsection.

4.4.3 Results

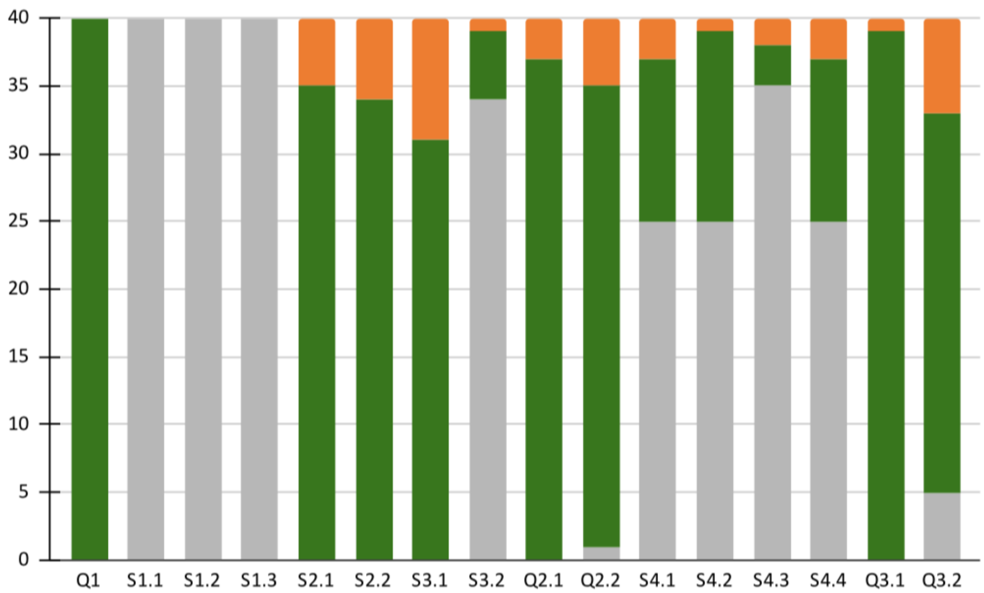
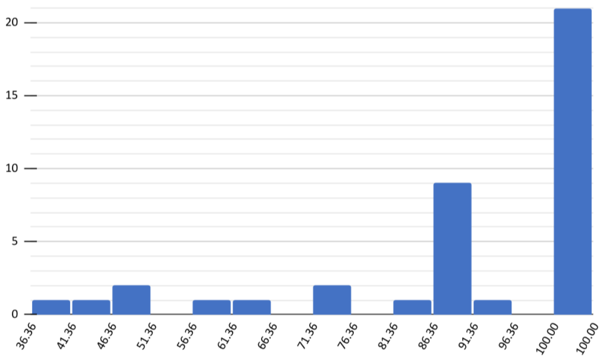


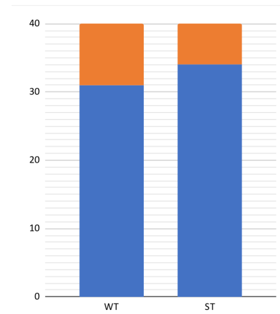
Figure 4.18: The frequency of checklist questions ['Green' means 'Achieved', 'Orange' means 'Developing' and 'Grey' means 'Not applicable']

Inter-grader agreement

The Cohen-Kappa score, which was calculated based on the grading of five sample recordings by two independent reviewers, was 0.32. This value implies that the reviewers had 96% agreement, which is marked as 'fair' on the Cohen-Kappa scale. Although the level of agreement is fair, the score was a bit low as the agreement



(a) Percentage of fidelity scores



(b) Number of students (in blue) achieving "warm and encouraging tone" and "structure"

Figure 4.19: Evaluation of student's recordings by reviewers

between the reviewers differed greatly on the point of 'warm and encouraging tone.' Out of all the points in the skills checklist, this particular category is a relative one. It might be greatly impacted by the individual reviewer's culture, mindset, and experience level. A better agreement score can be achieved by increasing the sample number and ensuring a common training session.

Frequency of questions answered in skills checklist

Figure 4.18 shows the frequency of questions and skills which were either 'achieved' (Green) or 'developing' (Orange) by the trainee facilitators during this study. The grey ones mean that those skills or questions were not applicable in that particular session. The questions and skills plotted in the graph are in a sequence following the Figure 4.17. From the graph, we see that the question at the very beginning that invites to make an action plan has been used in 100% cases. The behavioral menu at the beginning was not necessary at all in this scenario. However, the behavioral menu has been moderately used in problem-solving for low confidence levels.

Fidelity Test Score

Figure 4.19a shows the percentage scores achieved by the students after the grading performed by the reviewers. 85% students were able to pass the fidelity test, and more than 50% students passed the test with a 100% score. 78% students were able to maintain warmth and encouraging tone during the session, and 85% could maintain the core structure of intervention.

4.4.4 Discussion

The overall results and insights gained from this study are very promising and show that our CAP tool can support training for facilitators. The students performed exceptionally well given the minimal training period with such an intervention method. Although some students failed the fidelity test, that might be because the sessions were part of a voluntary and ungraded assignment. We can provide better inference by conducting the study in a cohort of professionals.

CHAPTER 5
Conclusion and Future Work

5.1 Broader Impact of this dissertation

This dissertation addresses a number of challenges in preventing vision loss from diabetes. We can summarize the contributions in the following way:

- Although there is a number of tools, e.g., Epic, Cerner etc. for telemedicine in clinical settings, there is none compatible for community settings. Our developed mTOCS system is a novel contribution that can be extended for any other community healthcare and reduce healthcare disparities among minority communities.
- Extensive research has been done in automated grading of retinal screening; however, very little has been done to make the models explainable. Our exploratory study in this area is very contemporary and has several future directions of extension.
- Our developed CAP tool is a state-of-the-art contribution in the field of Behavioral Science and Physical Therapy clinical care. It can be extended to other interventions including, stress management, nutrition management, pain management, etc.

5.2 Future Work

Future work can integrate an automated retinal screening model into the community telemedicine framework and thus, develop a standalone framework by also incorporating the CAP tool for overall health management of diabetic population in underprivileged communities. Upon completing the standalone framework, we can design a controlled clinical trial by recruiting a specific number of participants who have already had diabetes for a prolonged period. The participants would have retinal screening through our system and make an action plan to manage their diabetes. While selecting the participants, we can invite people already existing in our mTOCS database. We would include people whose data have not been used for training, which would make our data more diverse by including a random proportion of disease and normal eyes. We can evaluate the accuracy of the screening by comparing it with the grading by ophthalmologists. For the efficacy of the action plan, we can follow up with the participants after one month. We would conduct a survey to identify if the action planning process has helped them manage their diabetes well.

In the development of the explainable retinal screening model, we have implemented only one of many techniques for achieving explainability, and there are a lot more opportunities to extend this work in the future. Zhang et al. [67] have proposed a method to modify a traditional convolutional neural network (CNN) into an interpretable CNN to clarify knowledge representations in high conv-layers of the CNN. In an interpretable CNN, each filter in a high conv-layer represents a specific object part. The interpretable CNNs use the same training data as ordinary CNNs without annotating object parts or textures for supervision. The interpretable CNN automatically assigns each filter in a high conv-layer with an object part during the learning process. The explicit knowledge representation in an interpretable CNN can help people understand the logic inside a CNN, i.e., what patterns are memorized by CNN for prediction. Through experiments, the author has shown that filters in an interpretable CNN are more semantically meaningful than those in a traditional CNN. The authors have published their code for the newly proposed CNN architecture as open-source. Also, the validations of their results are published on the VGG-16 and AlexNet architecture. This work by Zhang et al. can be extended on DenseNet-121, as used in this dissertation work, and thus, can be evaluated for retinal screening in particular. The performance of the model can be measured by comparing the accuracy of the intrinsic interpretable model with our current model by keeping the dataset and all hyper-parameters constant during the training phase.

Our CAP tool has excellent potential as data architecture and electronic framework to apply the behavioral process of action planning to health specialties where these behavioral strategies have not traditionally been employed, such as rehabilitation. CAP is charting new territory and has the potential for development and assessment in future areas. Future work will include testing of CAP to assess whether its use better facilitates successful plan development and health behavior change relative to usual clinical care. A predictive validation study can be conducted to identify which data elements within CAP predict the successful completion of action plans. With respect to educational research, CAP can be assessed as a means to train new facilitators to become proficient in BAP. Over a long duration, we can also measure how CAP improves patient-relevant outcomes when implemented in the clinical episode of care and how CAP impacts modern medicine.

In summary, this dissertation work has multiple directions of future works which would be highly impactful and significant in preventing vision loss from diabetes.

BIBLIOGRAPHY

[1] (2010) Messidor-2 dataset. [Online]. Available: http://www.adcis.net/en/third-party/messidor2/

[2] (2015) Diabetic retinopathy detection. [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection/data

[3] (2019) Aptos 2019 blindness detection. [Online]. Available: https://www.kaggle.com/c/aptos2019-blindness-detection/data

[4] M. D. Abràmoff, J. C. Folk, D. P. Han, J. D. Walker, D. F. Williams, S. R. Russell, P. Massin, B. Cochener, P. Gain, L. Tang *et al.*, "Automated analysis of retinal images for detection of referable diabetic retinopathy," *JAMA ophthalmology*, vol. 131, no. 3, pp. 351–357, 2013.

[5] M. D. Abràmoff, M. Niemeijer, M. S. Suttorp-Schulten, M. A. Viergever, S. R. Russell, and B. Van Ginneken, "Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes," *Diabetes care*, vol. 31, no. 2, pp. 193–198, 2008.

[6] M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Investigative ophthalmology & visual science*, vol. 57, no. 13, pp. 5200–5206, 2016.

[7] E. Aharonovich, P. C. Amrhein, A. Bisaga, E. V. Nunes, and D. S. Hasin, "Cognition, commitment language, and behavioral change among cocaine-dependent patients." *Psychology of Addictive Behaviors*, vol. 22, no. 4, p. 557, 2008.

[8] A. D. Association, "Eye complications — ada," 2019, retrieved September 19, 2019 from https://www.diabetes.org/diabetes/complications/eye-complications.

[9] A. Barsegian, B. Kotlyar, J. Lee, M. O. Salifu, and S. I. McFarlane, "Diabetic retinopathy: Focus on minority populations," *International journal of clinical endocrinology and metabolism*, vol. 3, no. 1, p. 034, 2017.

[10] M. Bhaskaranand, C. Ramachandra, S. Bhat, J. Cuadros, M. G. Nittala, S. R. Sadda, and K. Solanki, "The value of automated diabetic retinopathy screening with the eyeart system: A study of more than 100,000 consecutive encounters from people with diabetes," *Diabetes technology & therapeutics*, vol. 21, no. 11, pp. 635–643, 2019.

[11] T. Bodenheimer and M. A. Handley, "Goal-setting for behavior change in primary care: an exploration and status report," *Patient education and counseling*, vol. 76, no. 2, pp. 174–180, 2009.

[12] N. M. Bolster, M. E. Giardini, and A. Bastawrous, “The diabetic retinopathy screening workflow: potential for smartphone imaging,” *Journal of diabetes science and technology*, vol. 10, no. 2, pp. 318–324, 2016.

[13] Cerner, 2020, last accessed on Oct 2020. [Online]. Available: http://www.cerner.com/

[14] N. Chaturvedi, M. Porta, R. Klein, T. Orchard, J. Fuller, H. H. Parving, R. Bilous, A. K. Sjølie, D. P. S. Group *et al.*, “Effect of candesartan on prevention (direct-prevent 1) and progression (direct-protect 1) of retinopathy in type 1 diabetes: randomised, placebo-controlled trials,” *The Lancet*, vol. 372, no. 9647, pp. 1394–1402, 2008.

[15] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, “Disease prediction by machine learning over big data from healthcare communities,” *Ieee Access*, vol. 5, pp. 8869–8879, 2017.

[16] N. Cho, J. Shaw, S. Karuranga, Y. Huang, J. da Rocha Fernandes, A. Ohlrogge, and B. Malanda, “Idf diabetes atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045,” *Diabetes research and clinical practice*, vol. 138, pp. 271–281, 2018.

[17] R. B. Cialdini, *Influence: Science and practice*, 2009, vol. 4.

[18] P. Costa, T. Araújo, G. Aresta, A. Galdran, A. M. Mendonça, A. Smailagic, and A. Campilho, “Eyewes: weakly supervised pre-trained convolutional neural networks for diabetic retinopathy detection,” in *2019 16th International Conference on Machine Vision Applications (MVA)*. IEEE, 2019, pp. 1–6.

[19] J. Cuadros and G. Bresnick, “Eyepacs: an adaptable telemedicine system for diabetic retinopathy screening,” *Journal of diabetes science and technology*, vol. 3, no. 3, pp. 509–516, 2009.

[20] K. P. Derose and D. W. Baker, “Limited english proficiency and latinos’ use of physician services,” *Medical Care Research and Review*, vol. 57, no. 1, pp. 76–91, 2000.

[21] M. Du, N. Liu, and X. Hu, “Techniques for interpretable machine learning,” *Communications of the ACM*, vol. 63, no. 1, pp. 68–77, 2019.

[22] Epic, 2020, last accessed on Oct 2020. [Online]. Available: http://www.epic.com/

[23] J. Fischer, T. Otto, F. Delori, L. Pace, and G. Staurenghi, “Scanning laser ophthalmoscopy (slo),” in *High Resolution Imaging in Microscopy and Ophthalmology*. Springer, 2019, pp. 35–57.

[24] C. for Disease Control and Prevention, “Press release — cdc,” 2010, retrieved October 20, 2020 from https://www.cdc.gov/media/pressrel/2010/r101022.html.

[25] ——, “Age-adjusted percentage of adults aged 18 years or older with

diagnosed diabetes who were physically inactive,” 2020, retrieved Oct 22, 2020 from https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.Pdf.

[26] ——, “Vision and eye health surveillance system — vision health initiative (vhi) — cdc,” 2020, retrieved Oct 22, 2020 from https://www.cdc.gov/visionhealth/vehss/index.html.

[27] T. H. J. K. F. Foundation, “Key facts about the uninsured population,” February 2019, retrieved September 19, 2019 from https://www.kff.org/uninsured/fact-sheet/key-facts-about-the-uninsured-population/.

[28] R. Gargeya and T. Leng, “Automated identification of diabetic retinopathy using deep learning,” *Ophthalmology*, vol. 124, no. 7, pp. 962–969, 2017.

[29] N. M. Gell, D. E. Rosenberg, G. Demiris, A. Z. LaCroix, and K. V. Patel, “Patterns of technology use among older adults with and without disabilities,” *The Gerontologist*, vol. 55, no. 3, pp. 412–421, 2015.

[30] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[31] D. Gutnick, K. Reims, C. Davis, H. Gainforth, M. Jay, and S. Cole, “Brief action planning to facilitate behavior change and support patient self-management,” *JCOM*, vol. 21, no. 1, pp. 17–29, 2014.

[32] J. Hadley, E. P. Steinberg, and J. Feder, “Comparison of uninsured and privately insured hospital patients: condition on admission, resource use, and outcome,” *Jama*, vol. 265, no. 3, pp. 374–379, 1991.

[33] K. Harris Nwanyanwu, H. K. Grossetta Nardini, G. Shaughness, M. Nunez-Smith, and P.-A. Newman-Casey, “Systematic review of community-engaged research in ophthalmology,” *Expert review of ophthalmology*, vol. 12, no. 3, pp. 233–241, 2017.

[34] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, “Metrics for explainable ai: Challenges and prospects,” *arXiv preprint arXiv:1812.04608*, 2018.

[35] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[36] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, “Artificial intelligence in healthcare: past, present and future,” *Stroke and vascular neurology*, vol. 2, no. 4, pp. 230–243, 2017.

[37] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, “Mimic-iii, a freely accessible critical care database,” *Scientific data*, vol. 3, p. 160035, 2016.

[38] B. Knowles and V. L. Hanson, "The wisdom of older technology (non) users," *Communications of the ACM*, vol. 61, no. 3, pp. 72–77, 2018.

[39] D. Ledo, S. Houben, J. Vermeulen, N. Marquardt, L. Oehlberg, and S. Greenberg, "Evaluation strategies for hci toolkit research," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 36.

[40] K. Lorig, H. Holman, and D. Sobel, *Living a healthy life with chronic conditions: self-management of heart disease, arthritis, diabetes, depression, asthma, bronchitis, emphysema and other physical and mental health conditions*. Bull Publishing Company, 2012.

[41] N. Menachemi and T. H. Collum, "Benefits and drawbacks of electronic health record systems," *Risk management and healthcare policy*, vol. 4, p. 47, 2011.

[42] W. R. Miller and S. Rollnick, *Motivational interviewing: Helping people change*. Guilford press, 2012.

[43] S. Mills, "Electronic health records and use of clinical decision support," *Critical Care Nursing Clinics*, vol. 31, no. 2, pp. 125–131, 2019.

[44] N. Mora, J. H. Kempen, and L. Sobrin, "Diabetic retinopathy in hispanics: A perspective on disease burden," *American journal of ophthalmology*, vol. 196, pp. xviii–xxiv, 2018.

[45] A. Nguyen, A. Dosovitskiy, J. Yosinski, T. Brox, and J. Clune, "Synthesizing the preferred inputs for neurons in neural networks via deep generator networks," 2016.

[46] W. D. of Health Services, "Hispanic/latinos in wisconsin: Health facts," September 2018, retrieved September 19, 2019 from https://www.dhs.wisconsin.gov/minority-health/population/hispanlatino-pophealth.htm.

[47] K. O'Leary, S. M. Schueller, J. O. Wobbrock, and W. Pratt, ""suddenly, we got to become therapists for each other": Designing peer support chats for mental health," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 331.

[48] W. H. Organization, "Diabetes," 2019, retrieved September 19, 2019 from https://www.who.int/health-topics/diabetes.

[49] A. N. Ortega, H. Fang, V. H. Perez, J. A. Rizzo, O. Carter-Pokras, S. P. Wallace, and L. Gelberg, "Health care access, use of services, and experiences among undocumented mexicans and other latinos," *Archives of internal medicine*, vol. 167, no. 21, pp. 2354–2360, 2007.

[50] R. E. Petty, P. Briñol, L. R. Fabrigar, and D. T. Wegener, "Attitude structure and change," *Advanced Social Psychology: The State of the Science*, p. 117, 2019.

[51] K. Reims, D. Gutnick, C. Davis, S. Cole, and B. Hope, “Brief action planning,” *Centre for Collaboration, motivation and innovation*, 2013.

[52] M. T. Ribeiro, S. Singh, and C. Guestrin, ““why should i trust you?” explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

[53] R. Schwarzer, “Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors,” *Applied psychology*, vol. 57, no. 1, pp. 1–29, 2008.

[54] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, “Grad-cam: Why did you say that?” 2016.

[55] P. D. Soderlund, “Effectiveness of motivational interviewing for improving physical activity self-management for adults with type 2 diabetes: A review,” *Chronic Illness*, vol. 14, no. 1, pp. 54–68, 2018.

[56] K. Solanki, C. Ramachandra, S. Bhat, M. Bhaskaranand, M. G. Nittala, and S. R. Sadda, “Eyeart: automated, high-throughput, image analysis for diabetic retinopathy screening,” *Investigative Ophthalmology & Visual Science*, vol. 56, no. 7, pp. 1429–1429, 2015.

[57] A. Sopharak, B. Uyyanonvara, S. Barman, and T. H. Williamson, “Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods,” *Computerized medical imaging and graphics*, vol. 32, no. 8, pp. 720–727, 2008.

[58] D. S. W. Ting, C. Y.-L. Cheung, G. Lim, G. S. W. Tan, N. D. Quang, A. Gan, H. Hamzah, R. Garcia-Franco, I. Y. San Yeo, S. Y. Lee *et al.*, “Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes,” *Jama*, vol. 318, no. 22, pp. 2211–2223, 2017.

[59] A. Tufail, C. Rudisill, C. Egan, V. V. Kapetanakis, S. Salas-Vega, C. G. Owen, A. Lee, V. Louw, J. Anderson, G. Liew *et al.*, “Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders,” *Ophthalmology*, vol. 124, no. 3, pp. 343–351, 2017.

[60] A. A. Van Der Heijden, M. D. Abramoff, F. Verbraak, M. V. van Hecke, A. Liem, and G. Nijpels, “Validation of automated screening for referable diabetic retinopathy with the idx-dr device in the hoorn diabetes care system,” *Acta ophthalmologica*, vol. 96, no. 1, pp. 63–68, 2018.

[61] G. A. G. Vimala and S. K. Mohideen, “An economic system for screening of diabetic retinopathy using fundus images,” *OnLine Journal of Biological Sciences*, vol. 14, no. 4, p. 254, 2014.

[62] T. Wan, H. Jun, P. W. Hui ZHANG, and H. Hua, “Kappa coefficient: a

popular measure of rater agreement," *Shanghai archives of psychiatry*, vol. 27, no. 1, p. 62, 2015.

[63] X.-N. Wang, L. Dai, S.-T. Li, H.-Y. Kong, B. Sheng, and Q. Wu, "Automatic grading system for diabetic retinopathy diagnosis using deep learning artificial intelligence software," *Current Eye Research*, pp. 1–6, 2020.

[64] J. W. Yau, S. L. Rogers, R. Kawasaki, E. L. Lamoureux, J. W. Kowalski, T. Bek, S.-J. Chen, J. M. Dekker, A. Fletcher, J. Grauslund *et al.*, "Global prevalence and major risk factors of diabetic retinopathy," *Diabetes care*, vol. 35, no. 3, pp. 556–564, 2012.

[65] A. Zambelli-Weiner, J. E. Crews, and D. S. Friedman, "Disparities in adult vision health in the united states," *American journal of ophthalmology*, vol. 154, no. 6, pp. S23–S30, 2012.

[66] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," 2013.

[67] Q. Zhang, Y. Nian Wu, and S.-C. Zhu, "Interpretable convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8827–8836.

[68] X. Zhang, M. F. Cotch, A. Ryskulova, S. A. Primo, P. Nair, C.-F. Chou, L. S. Geiss, L. E. Barker, A. F. Elliott, J. E. Crews *et al.*, "Vision health disparities in the united states by race/ethnicity, education, and economic status: findings from two nationally representative surveys," *American journal of ophthalmology*, vol. 154, no. 6, pp. S53–S62, 2012.

[69] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," 2015.

[70] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," *Electronics*, vol. 10, no. 5, p. 593, 2021.