The Usability and Trustworthiness of Medical Eye Images

Daniel Diethei
University of Bremen
Bremen, Germany
diethei@uni-bremen.de

Ashley Colley
University of Lapland
Rovaniemi, Finland
ashleycolley@gmail.com

Lisa Dannenberg
University of Bremen
Bremen, Germany
lisadannenberg2@gmail.com

Muhammad Fawad Jawaid Malik
University of Bremen
Bremen, Germany
malikm@uni-bremen.de

Johannes Schöning
University of Bremen
Bremen, Germany
schoening@uni-bremen.de

Abstract—The majority of blindness is preventable, and is located in developing countries. While mHealth applications for retinal imaging in combination with affordable smartphone lens adaptors are a step towards better eye care access, the expert knowledge and additional hardware needed are often unavailable in developing countries. Eye screening apps without lens adaptors exist, but we do not know much about the experience of guiding users to take medical eye images. Additionally, when an AI based diagnosis is provided, trust plays an important role in ensuring in the adoption. This work addresses factors that impact the usability and trustworthiness dimensions of mHealth applications. We present the design, development and evaluation of *EyeGuide*, a mobile app that assists users in taking medical eye images using only their smartphone camera. In a study (n=28) we observed that users of an interactive tutorial captured images faster compared to audible tone based guidance. In a second study (n=40) we found out that providing disease-specific background information was the most effective factor to increase trustworthiness in the AI based diagnosis. Application areas of *EyeGuide* are AI based disease detection and telemedicine examinations.

Index Terms—mHealth, eHealth, Smartphones, Mobile devices, AI Diagnosis, Telemedicine, Eye diseases

I. Introduction & Motivation

According to the World Health Organization (WHO), 65.2 million people have cataract, a cloudiness in the lens of the eye, leading to increasingly blurred vision [1]. Of the visually impaired and blind, 90% live in developing countries and 80% of visual impairments are preventable [2]. In many countries, the poorest households are more likely to have access to mobile phones than to toilets or clean water [3]. With this background, mobile health (mHealth) technologies that can support the early detection of causes of preventable blindness should be explored with high priority. Retinal screening programs for common eye diseases can provide early detection of chronic eye diseases, but come at a high cost, e.g., a stationary slit lamp camera for traditional ophthalmology costs about $15.000 [4]. One step to increase the affordability of retinal screening is to use a shared community smartphone fitted with a lens adaptor, e.g. Peek Retina ($240) or D-Eye ($435). Such
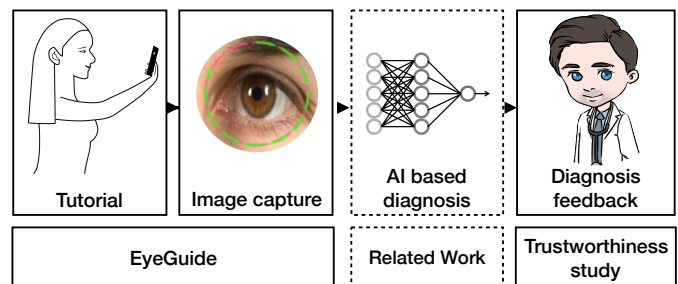

Fig. 1. Smartphone based eye disease diagnosis app. The scope of this paper covers (1) the tutorial and image capture phases and (2) the diagnosis feedback, focusing on the trustworthiness of the diagnosis. The AI based diagnosis functionality is not in the scope of this paper.

solutions can be applied by clinicians and nurses in areas with limited healthcare facilities and have been shown to perform well in disease detection. [4]

Whilst the increasing availability of smartphones has resulted in a growing market for mHealth applications, e.g. in the areas of respiratory diseases [5] and dermatology [6], there is little evidence of positive clinical outcomes [7]. Studies have suggested that users require consumer-friendly devices and apps that are self-reinforcing and enjoyable to use [8]. Reporting on experiences in the UK's National Health Service, [9] highlight ease of use of mHealth apps as a deciding factor in their adoption. A further barrier to the adoption of mHealth services is their perceived trustworthiness, e.g. in the competence of the provider, privacy or security of the service [10, 11, 12].

To address the issues mentioned above, we developed a user-friendly smartphone app for eye disease detection. In this paper, we focus specifically on the phases of image capture and the presentation of the diagnosis to the user (Figure 1). To account for a wide range of individual's living circumstances, we target a solution enabling self-taking of medical eye images, and focus on the use of the smartphone's higher quality rear camera. Whilst there is much prior work

on automated diagnosis of eye disease from images, e.g. using deep learning [13, 14, 15, 16], the effectiveness of this is limited if the source images for analysis are of poor quality.

The contribution of this work is twofold as we connect the fields of usability and trustworthiness in the context of mHealth. As suggested by others [17, 18, 19], the usability of mHealth applications impacts trustworthiness, i.e. users find applications that consider usability principles to be more trustworthy. Therefore, we describe the design, development and evaluation of the multi-modal *EyeGuide* app, that guides users in the capture of eye images suitable for detection of eye diseases such as cataracts. To support the task, the app includes a tutorial phase and uses auditory and visual guidance to direct the user to position the camera correctly. As a second contribution, we present a study exploring different approaches to increase the perceived trustworthiness of the eye disease diagnosis app.

From our evaluation, we report that the interactive tutorial had advantages in terms of usability of quality of taken images compared to the auditory guidance. A review of the taken images by an ophthalmologist indicated that the images taken with our app are suitable for diagnosing diseases in the conjunctiva and cornea. The most effective way to build trustworthiness in the app was through increasing users' knowledge level about the disease.

II. RELATED WORK

Our work in the area of intelligent user interfaces builds on three areas of prior work on (1) smartphone based mHealth for eye diseases, (2) guiding image capture, and (3) trustworthiness in mHealth. Firstly, we introduce existing mHealth applications addressing eye-related diseases. We then present works with different approaches to guide users when taking photographs, e.g. ensuring the subject is correctly located in the frame. Finally, we give a short overview of issues surrounding the perceived trustworthiness of mHealth applications.

A. Smartphone-Based Eye Disease Diagnosis

Prior work has presented smartphone apps to diagnose disease based on eye-images, e.g. for cataract detection [20, 21], to identify high cholesterol levels [22, 23], to diagnose concussions [24] and for glaucoma screening [23, 25]. Akil and Elloumi [26] present a meta paper, investigating the image quality and diagnosis performance achieved in eight prior works using smartphones equipped with additional lenses for retinal examination. Most recently deep learning has been presented as an approach to identify eye diseases [13, 14, 15, 16]. For example, Wei et al. [14] presented a deep learning based smartphone app to identify retinal abnormalities. The system gives simple real-time textual feedback 'normal / disease detected', requires the use of an additional D-Eye lens fitted to the smartphone and cannot be self-administered. Kim et al. [27] developed and evaluated an automated smartphone-based system for retinal disease screening using a 3D printed housing for acquiring a series of seven images stitched together to a widefield retinal montage. Munson et al. [28] developed a free

smartphone app able to identify 'white eye' and subsequent eye disorders based on casual photographs of small children. The authors reported the app correctly diagnosed 80% of children with eye disorders. EyeGuide builds upon work by Diethei and Schöning [29] who explored an agent-based tutorial for eye image guidance. In this paper, we present a study to assess the usability of auditory guidance and an interactive tutorial to take eye images without the need of additional lenses or other hardware.

B. Guiding Image Capture

Ensuring correctly positioned and high quality images of the eye is a critical element in the performance of the following diagnosis phases, either manually by a doctor or through AI based solutions. As the target of our approach is towards the creation of a self-administered eye examination tool, using the device's higher quality rear camera, solutions to guide the user in image capture are needed. Similar challenges have been addressed in mHealth apps for dermatology. For example, in the Skinvision app [30] the user "waves" the smartphone over the skin lesion to be investigated, and the app automatically selects when to take the image for analysis. This unguided approach could potentially be improved on with audio feedback [31], e.g., similar to that used for navigation [32]. A summary of approaches to integration non-speech sounds to visual interfaces has been presented by [33], suggesting the use of earcons [34, 35], sounds with dynamic timber, pitch and rhythm, as suitable for localization tasks. To guide users to take better portrait photos with smartphones, [36] demonstrated a solution using verbal guidance e.g. "please move to the left", to which [37] added vibration modality.

C. Trustworthiness

The importance of the user interface in perceived trustworthiness has been reported by several works, e.g., [17, 18, 19]. Kim and Moon [19] have shown that the user interface design factors impact customer confidence. Roy et al. [17] found out that the usability factors ease of navigation, ease of learning, perception and support were associated with trustworthiness. In the field of eHealth, Vo et al. [38] highlighted trustworthiness, appropriateness, personalization, and accessibility as the main weaknesses in current mHealth apps. Fruhling and Lee [39] developed a usability model for the consumer's perception of trust in eHealth services, noting that websites with a higher usability were likely to be viewed as more trustworthy. Going beyond usability, others have highlighted the role of visual aesthetics in the design of mHealth apps in their perceived credibility [40]. However, mHealth app designers do not have free hands, the requirements for apps to meet regulatory requirements can impact the design, and consequentially affect the perceived trustworthiness the app [41].

D. Summary of Prior Work

From the related work we learn that, whilst much research effort has been spent addressing automated eye-disease diagnosis from images, little research has addressed the activities
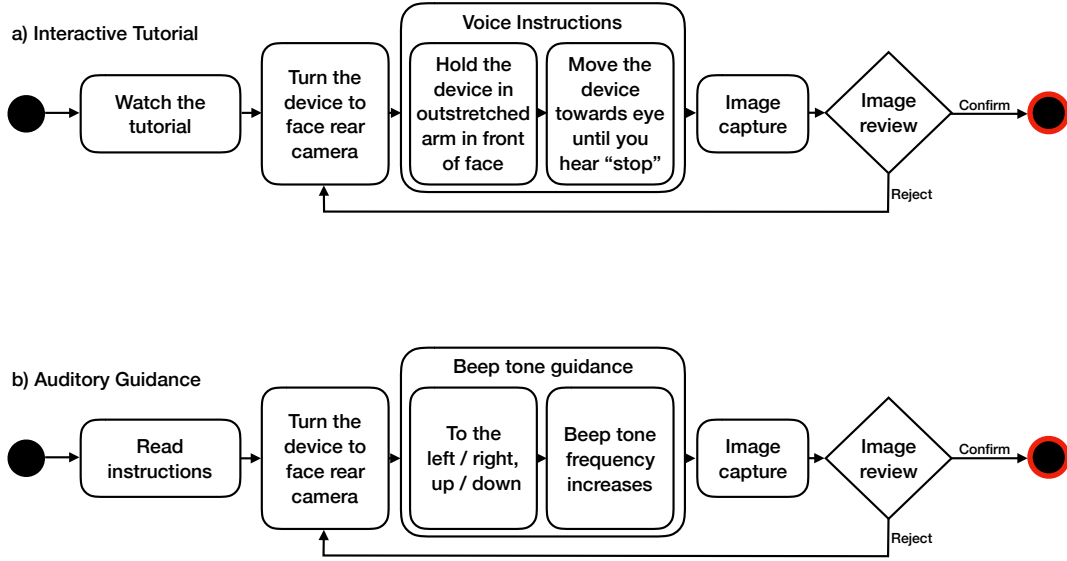
Fig. 2. Flow diagram of the two study conditions. Participants in the interactive tutorial group first watched a tutorial demonstrating the procedure to take eye images and then followed voice instructions. In the auditory guidance group, participants only read brief instructions to take an image of the iris and were then supported through voice and beep tone guidance.

preceding or following this in an operational mHealth solution. With a focus on these phases we, (1) demonstrate and evaluate two solutions using audio signals and an interactive tutorial to guide users to take optimal medical eye images (2) present design alternatives aiming to improve the perceived trustworthiness of mHealth smartphone apps. In particular our approach follows Akter et al.'s [42] dimensions of trustworthiness; ability, benevolence, integrity and predictability.

III. EYEGUIDE

The *EyeGuide* app guides the user to take high quality images of their own eye, using the rear camera of a smartphone. These images are suitable for analysis by a clinician or through an automated AI process. As in this configuration the smartphone's screen is facing away from the user, the user must be guided to position the camera to fully capture the eye.

A. Concept

To understand the issues related to clinical eye photography on a smartphone, we first instructed four individuals to acquire photos of their iris with their smartphone in a prestudy. The participants used the native camera application with both the front and rear camera. An analysis of the resulting images revealed that lighting and device positioning seemed to be the predominant causes of blurry images or images without a fully visible iris. The front camera did not provide sufficient image quality for our purpose of detecting diseases due to the lack of an autofocus functionality and the lower image resolution. Although the front cameras of the latest high-end smartphones are of higher resolution than those used in the test, considering our target populations in developing countries, the use of older, lower-cost devices with limited front camera resolution is more representative.

Based on previous work by Diethei and Schöning [29] and the cognitive theory of multimedia learning [43], two variants were implemented:

- Interactive tutorial (referred to as IT)
- Reading instructions followed by voice and audio tone based guidance (referred to as 'auditory guidance'; AG)

The procedures of the two app variants are illustrated in Figure 2.

In the interactive tutorial (IT) case, the process of positioning the phone is first demonstrated with an on-screen animation (Figure 3). After viewing the complete animation, the user is guided by voice instructions to complete each of the steps in turn. When the camera is in the right position the user is instructed to 'stop' and, after a countdown of three seconds, the camera flash light is activated to illuminate the eye evenly
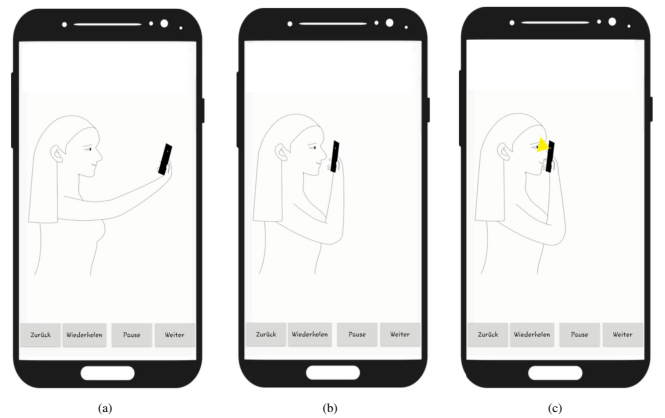

Fig. 3. Tutorial to take the eye photos. Left (a): Arm outstretched in front of face. Center (b): Moving the phone closer to the eye until the word "stop" is played back. Right (c): Countdown and image acquisition with flash.

and the photo is taken. In the auditory guidance (AG) case, voice guidance is used for coarse positioning, e.g. 'to the right', 'up'. When the eye is recognised in the middle of the screen, the audio output consists of a sequence of short tones, which become more frequent with better camera positioning (cf. Geiger counter metaphor, [32]). As soon as the camera is in the correct position, the audio tone becomes continuous, the camera flash light is illuminated and the eye photo is automatically taken. In both IT and AG cases successful acquisition is indicated through a camera shutter sound and vibration feedback. At the end of the process, an image review screen is shown to the user with the option to retake should the iris be blurry.

B. Implementation

EyeGuide is an Android app written in Java. To track the eye position we used the `openCV` library with the `Haar Feature-based Cascade Classifier for Object Detection` [44], which is an AI-based model for image recognition tasks. To find the optimal parameters for the eye detection (`detectMultiScale` method, parameters `scaleFactor` and `neighbors`), we took 50 sample images, 28 with eyes and 22 without, with the rear camera of a Samsung Galaxy S7. The images without eyes showed other parts of the face around the eyes. We then analyzed the classifications (correct, false positive, false negative) for combinations of the parameters. Finally, we chose the parameters with the highest accuracy and speed.

For the interactive tutorial (IT) case, graphics and animations were created using standard graphic design applications. To design the tutorial and the virtual agent with a focus on teaching motor skills, we adhered to the cognitive theory of multimedia learning [43], in particular the pre-training and segmenting principle. The voice instructions were recorded by author LD. To give feedback about the position of the eye in the auditory guidance (AG) case, the camera view was divided into five virtual rectangles; one in the center of the frame and four around the center. If an eye was recognised in one of the outer rectangles, speech guidance was given using the Android Text-To-Speech engine, e.g. "to the right". The maximum frequency of the speech guidance was two seconds. When the eye was detected in the central rectangle, the guidance mechanism was changed to a sequence of short tones, which become more frequent with better camera positioning.

In both IT and AG cases, when the camera was optimally positioned, indicated either by 'stop' (IT) or a continuous tone (AG), a full-resolution photo was automatically saved to the phone. In the capture phase, guidance was based only on position, as the available computing power and lower quality camera stream did not support analysing the sharpness of the image in real time.

C. Evaluation

To evaluate the user experience of *EyeGuide* and its perceived diagnosis trustworthiness we conducted a user study. The study was conducted as a lab study and evaluated the
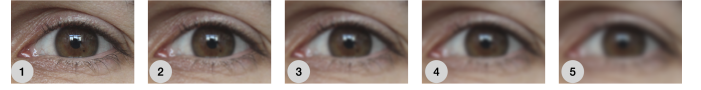

Fig. 4. To rate the image sharpness in the *EyeGuide* study, we used a series of images as a reference. From left to right: *completely sharp (1), sharp (2), slightly out of focus (3), blurry (4), completely blurry (5).*

two alternative designs, IT and AG, to enable comparison (Figure 2). We recruited 28 participants and assigned them to two groups (IT and AG) of 14. The IT group had a mean age of 23, while the AG group had an mean age of 22. Both groups consisted of six female and eight male participants. Five participants of the IT group and seven participants of the AG group wore glasses to correct their vision.

The study sessions were video recorded and all data collected during the study was stored anonymously. The participants were first introduced to the target of the study and its process and were then asked to sign a consent form. Afterwards, participants completed a demographic questionnaire and the TA-EG questionnaire [45], which measures technical affinity. Participants were encouraged to think-aloud during the test. The smartphone running the *EyeGuide* app was handed to the participants and they proceeded to follow the instructions given by the app. At the end of the test, participants responded to open questions about the experience of using *EyeGuide*.

D. Results

We firstly compared the number of attempts and time needed to take an acceptable eye image in both interactive tutorial (IT) and auditory guidance (AG) configurations. We conducted independent samples t-tests with α set to .05. Participants in the IT group (M_{tries} = 1.8, SD = .47) and in the AG group (M_{tries} = 1.3, SD = 1.19) did not differ in tries to capture the first eye, $t(26) = 1.46$, $p > .05$. However, the duration needed to take photos of both left and right eyes was significantly lower, $t(26) = -2.17$, $p < .05$, in the IT group ($M_{duration}$ = 151s, SD = 115.3s) than in the AG group ($M_{duration}$ = 247s, SD = 118s).

Author LD rated the sharpness of the eye images by comparing them to a set of pre-defined reference images (Figure 4). The number of images rated with each quality level are shown in Figure 5. A majority of images was either completely sharp or sharp in both groups (IT: 17, AG: 15) and only one image was completely blurry in each of the groups. No significant
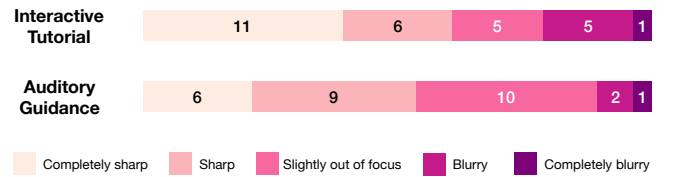

Fig. 5. Distribution of image sharpness ratings. Each participant (n = 28) took two images, one for each eye. There were no significant differences in image quality ($p > .05$) between the two groups (IT and AG).

difference between the two groups was apparent, $t(26) = .437$, $p > .05$, with mean ratings for the IT group $M_{imageQuality}$ = 3.75, SD = .87 and the AG group $M_{imageQuality}$ = 3.6, SD = .86. Due to a technical error in the interactive tutorial condition, the image capture was triggered at a larger distance compared to the auditory guidance condition. This might have improved the image quality since a lower device-eye distance can cause blurry photos.

In the interactive tutorial (IT) group, the on-screen agent instructed the participant to watch the following steps before putting them into practice. This instruction was viewed and confirmed by every participant. However, as the tutorial started, six out of 14 participants tried to follow the instructions by turning the smartphone around while the first step of tutorial was shown. Two of the six realized that it was an instructive section of the tutorial when they were asked to press the 'next' button, and turned the smartphone back around to complete watching the tutorial. The other four appeared to be confused on how to press the on-screen 'next' button while physically following the tutorial steps. Eight of the 14 participants in the IT group had at least one occasion where the word 'stop' was not played back, mostly due to to incorrect positioning of the device in front of the face. The participants then had to repeat the steps from the beginning. However, while the smartphone was being moved back to the start position, sometimes one of the eyes was recognized, which triggered the 'stop' instruction and resulting in participants capturing an image of their whole face. Other problems noted in the IT group included participants thinking they should capture both eyes at once, or not understanding that the photo would be taken automatically. Four out of

five participants wearing glasses removed their glasses to take the photo even though they were not asked to do so. Participants were instructed to start taking each photo with a fully outstretched arm, however, after going through the process once, most participants did not fully outstretch when taking subsequent photos.

In the auditory guided (AG) group, participants first read the instructions before taking any photos. After turning the smartphone around so they faced the rear camera, six of the 14 participants felt unsure about the correct distance from the eye to hold the smartphone. Four of the participants were not sure if they were required to take the photo themselves or if the photo would be taken automatically. In addition, eight participants were not able to interpret the meaning of the beeping sound. Two of the participants said that they do not know what to do when getting the voice instructions (to the left/right, up/down). Six participants turned the smartphone at least once to recheck the written instructions. In the AG group, four of the seven participants wearing glasses took their glasses off before taking the photos.

To summarise, the time to take images in the interactive tutorial (IT) version of *EyeGuide* was lower than in the audio guided (AG) version. The sequence of steps shown in the tutorial was clear to most study participants. However, we observed some usability issues, e.g. incorrectly carrying out physical instructions when participants had been instructed only to watch in the pre-training step.

IV. TRUSTWORTHINESS STUDY

Having evaluated different approaches for the tutorial and image capture phases, we also addressed the diagnosis feedback phase of smartphone-based eye disease detection. To

(a) *No trust-building (baseline)*. Ishihara [46] plates were shown in all design alternatives, except for the assessment diagnosis mismatch.

(b) *Assessment-diagnosis mismatch*. Participants provided speech samples that were not analyzed to allegedly detect color blindness.

(c) The *disclaimer* explained that the assessment performed in the study did not replace an examination by a professional ophthalmologist.

(d) In the *disease information* design variant, the participants were shown an explanation of possible causes for color blindness.
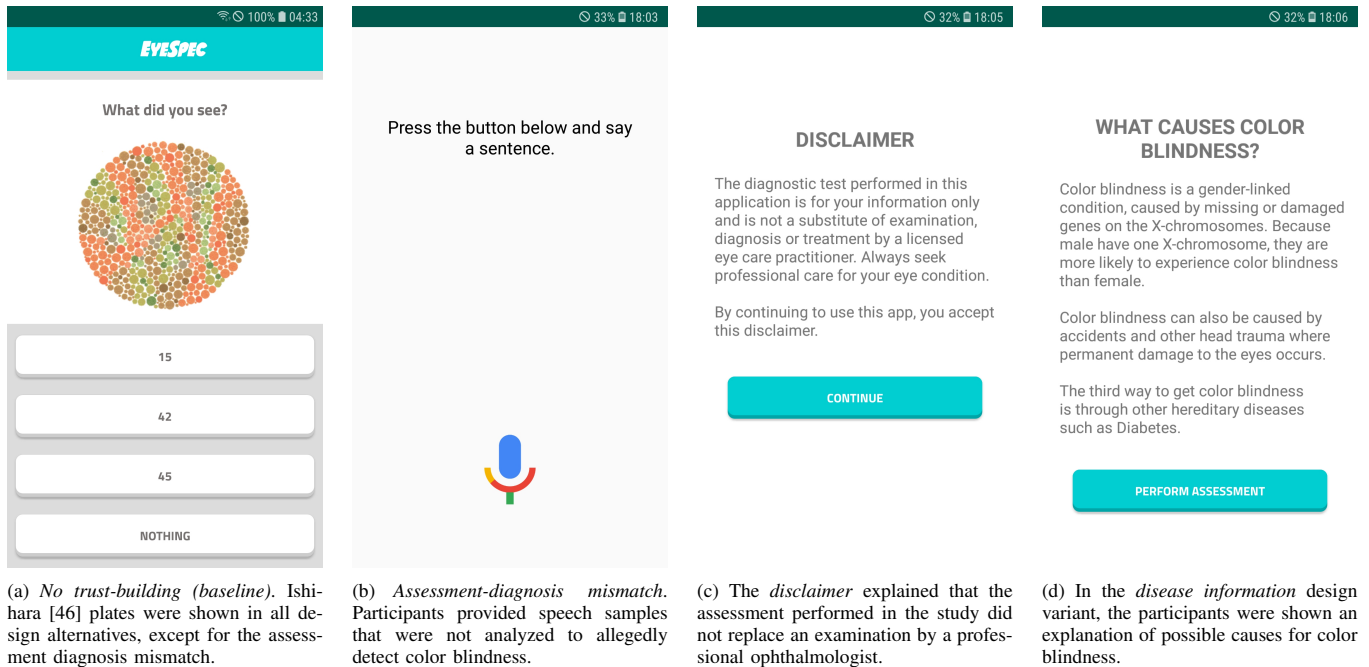
Fig. 6. The four design alternatives of the trustworthiness study.

gain insight into issues of trustworthiness in our targeted automated eye disease diagnosis app, we conducted a second experimental study.

A. Study Design

As the eye diseases we target with *EyeGuide* are rare in healthy Western populations, for the purpose of our study we decided to simulate the detection of the more common ailment of color blindness. We did not conduct any actual diagnoses and chose color blindness as an arbitrary common condition that is related to vision. As a study probe, we developed a separate Android app that included four alternative design approaches (Figure 6) to building trust. Each of the alternatives is mapped to one or two dimensions of the trustworthiness model. The users' interaction with the app was guided by an animated avatar, that represented the virtual ophthalmologist (Figure 7).

The four alternative designs included in our application were:

1) *No trust-building (baseline)*. This design represents the standard procedure for the assessment and diagnosis of color-blindness using the Ishihara test [46].
2) *Assessment-diagnosis mismatch*. This probe design examined if users trust the diagnosis even if there is a mismatch between the assessment and the diagnosis. Participants were asked to provide a speech sample to detect whether they have color blindness, even though there is no diagnostic relationship between these two modalities. This intervention is a negation to 'Predictability' from the trustworthiness model [42]. Predictability refers to the degree to which an application is expected to behave reliably by abiding to standard practices.
3) *Disclaimer*. With this design we aimed to find out whether informing the user that their data will not be used for wrong purposes and only with good intentions builds trust in the application. This targets the 'Integrity' and 'Benevolence' dimensions of the trustworthiness model. Benevolence refers to the good intentions of the application towards the user, whereas integrity refers to moral and ethical principles which an app should conform to. Therefore, to affirm both of these dimensions in the app, we display a disclaimer as well as terms and conditions at app start up.
4) *Disease information* This design addressed the 'Ability' dimension, which according to [42] is the most significant in the trustworthiness model. Ability refers to skills and competencies of the app that encourage the user to accept it, e.g. the application contains the desired knowledge about the task it performs. Addressing this, our design provided information in text form about color blindness to the user before actually starting the assessment.

B. Evaluation

To compare the trustworthiness of our four design variants we carried out a user study. The study followed a between-

Fig. 7. The virtual ophthalmologist avatar that leads through the color blindness test and diagnosis procedure.

subjects design with the independent variable *trust-building feature (no trust-building, assessment-diagnosis mismatch, disclaimer, disease information)* and the dependent variable *trustworthiness*. Participants were consecutively assigned to one of the four groups. Based on Akter et al. [42]'s trustworthiness model we developed the following hypotheses:

- H1: Giving background information about the condition results in the highest trustworthiness score.
- H2: The lowest trustworthiness is observed when there is a mismatch between assessment and diagnosis.
- H3: The highest benevolence score is achieved through providing a disclaimer as well as terms and conditions.
- H4: Integrity is highest through displaying a disclaimer and terms and conditions.
- H5: The ability score is highest when giving background information about the disease.

At the start of the session, we informed participants about the study both verbally and in writing. Participants then completed a consent form and a demographic questionnaire. After that, participants were given a smartphone containing the study app, which they opened and followed the instructions given by the app. At the end of the study session participants filled out a trustworthiness questionnaire and were given an opportunity to provide open feedback on the test process. During the study, the experimenter took notes. The trustworthiness questionnaire developed for the study was a version of that developed by [42], slightly modified for mHealth applications. The questionnaire consisted of multi-item scales with favorable psychometric properties. Each item in the questionnaire was measured in a structured arrangement on a 7-point Likert-type scale, ranging from 1 (strongly disagree) to 7 (strongly agree). There are 4 items for each of the dimensions of trustworthiness, and a final 4 questions directly investigating trust.

We recruited 40 participants (M_{Age} = 26.2, *SD* = 6.0), 18 of which were female and 22 male. Most of the participants were students of the University of Bremen.

C. Results

We first report findings on the overall trustworthiness score followed by the results referring to its sub-dimensions. To test the effect of the independent variable design alternative on the dependent variable trustworthiness, we conducted a one-way ANOVA with the significance level α = .05. For further analysis of the differences between groups, we used
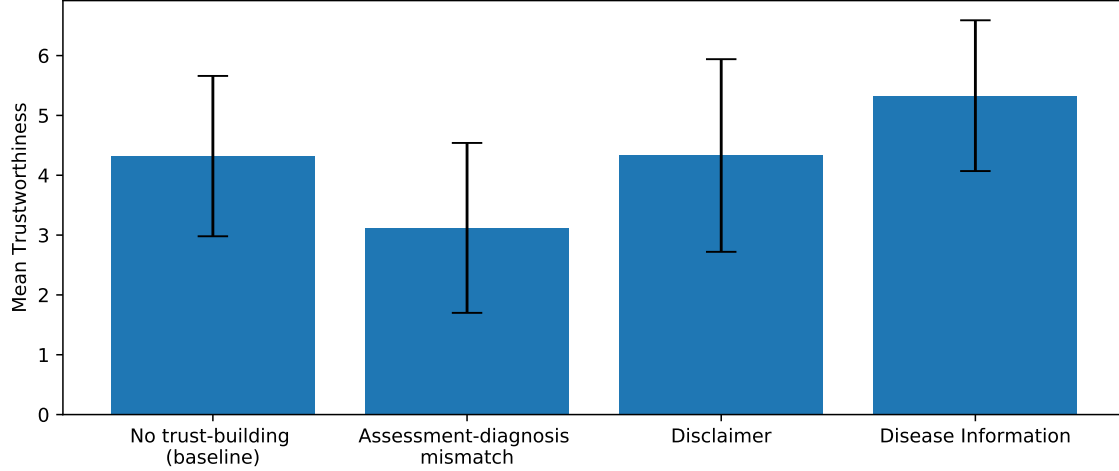
Fig. 8. Trustworthiness scores between design alternatives. Error bars represent standard deviation.

planned contrasts. When referring to individual participants, we indicate the design alternative and the participant number within the condition, e.g. P1.2 is participant 2 in the no trust-building condition (no trust-building = 1, assessment-diagnosis mismatch = 2, disclaimer = 3, disease information = 4).

As expected, we observed differences in trustworthiness between the design alternatives (Figure 8), $F(3,36) = 3.84$, $p < .05$, partial $\eta^2 = .24$. When provided with disease information ($M = 5.33$, $SD = 1.26$), participants reported higher trustworthiness compared to the other design alternatives, $t(36) = 2.68$, $p < .05$. Also, a mismatch between assessment and diagnosis led to the lowest trustworthiness, $t(36) = -2.86$, $p < .05$.

While the integrity dimension score was affected by the design alternative, $F(3,36) = 5.23$, $p < .05$, partial $\eta^2 = .30$, displaying a disclaimer did not lead to the highest integrity score, $t(36) = .79$, $p > .05$. The descriptive data indicate that providing disease information ($M = 4.63$, $SD = 1.29$) might increase integrity.

Similarly, the design alternative had an effect on the benevolence score, $F(3,36) = 3.08$, $p < .05$, partial $\eta^2 = .20$. However, contrary to our expectation, the disclaimer did not cause the highest benevolence, $t(36) = .54$, $p > .05$. Again, there are descriptive indications that disease information ($M = 5.55$, $SD = 1.49$) is more effective in increasing benevolence.

Lastly, there was a difference in ability, targeted through disease information, between the design alternatives $F(3,36) = 6.42$, $p < .001$, partial $\eta^2 = .35$. Consistent with our hypothesis, information about color-blindness led to higher ability, $t(36) = 3.18$, $p < .05$.

In the *no trust-building* group, seven out of ten participants complimented the app with statements like "nicely done" (P 2.10) and "fast" (2.9). When questioned whether they would trust the app or not only two of the participants said yes.

Participant 2.2 said she might have trusted it if she didn't know already that she is not color blind. Three out of ten participants in the *assessment-diagnosis mismatch* group were confused when presented with the results. Six participants questioned the diagnosis of color blindness based on speech samples. When shown the *disclaimer and terms and conditions*, seven out of ten participants read both screens. One individual skipped the terms and conditions text. All of the participants performed the Ishihara test correctly. One of the participants mentioned that they liked the app, especially the dialogue with the virtual ophthalmologist (P 3.6). Participant 3.5 commented that the app seems to have good intentions. On questioning them about their trust in the app, only participant 3.5 said that she trusted it, whilst participant 3.6 said she might trust it, but was a little hesitant.

We observed that nine out of ten participants in the *disease information* group read both the information screens, i.e. "what is color blindness?" and "what causes color blindness?". Participant 4.2 skipped the "what is color blindness?" screen. Six participants agreed that the app contains knowledge about the disease. Three participants rated the app with comments such as "user-friendly" (P 4.4) and "nice user interface" (4.10). Five out of ten participants said that they trusted the app. Participant 4.2 said she would trust the app if it was used in cooperation with a doctor.

V. DISCUSSION

Through two studies we firstly examined the usability of *EyeGuide*, a mHealth app to take medical eye photos and, secondly, identified design considerations to ensure trustworthiness of mHealth apps in general. While the evaluation of *EyeGuide* addressed the image capture phase of an automated eye disease diagnosis app, the trustworthiness study informed

on design approaches to increase trustworthiness in the diagnosis feedback phase.

A. EyeGuide

As some participants did not realise that the photo would be taken automatically and that they should only capture one eye at a time, not all tutorial instructions were understood correctly. Similar confusions about the meaning of the voice instructions (to the left/right, up/down) may have been due to a misunderstanding if the object to be moved was the smartphone or their head. On some occasions, participants repeated steps or activated functions by mistake because they were holding the smartphone in an incorrect position. Participants in both groups were confused when they were required to confirm the taken image.

Participants that wore glasses had similar problems with both concepts. All the photos taken by participants wearing glasses were of poor quality. Particularly, all those rated as completely out of focus were taken by glasses wearers and other images captured by glasses wearers were either blurry or slightly out of focus. To remedy this, participants removed their glasses - in hindsight this should have been instructed by the app.

A possible reason for the interactive tutorial group being faster in taking the eye images than the auditory guided tutorial group is that the auditory guided tutorial group took longer to find the correct device position. In contrast, the interactive tutorial group just had to follow the given instructions. Additionally, auditory guided participants often turned the smartphone around to check the instructions again, which added additional time. The mean time difference between the IT and AG group to complete the two images was on average about 100 seconds. In a real-world scenario, this has probably implications on the frustration levels of users, potentially leading to drop-outs in the AG concept.

In the interactive tutorial, some participants had trouble distinguishing between the sections of the tutorial where they were supposed to only watch the instructions and the sections where they had to actively carry out the instructions, e.g. actually turn around the phone. While the agent instructed the participants to first watch the steps before following them together, this was apparently not salient enough. We suggest a multi-modal indication of whether to watch only or carry out the instructions, e.g. a voice and text instruction.

In the auditory guidance group, many users did not recognize the Geiger counter metaphor, i.e. a beep tone with shorter intervals as they approached the target position.

Overall, we believe that the interactive tutorial concept had a better user experience. While some errors were detrimental to task completion and image quality, the concept of providing instructions that have to be carried out afterwards has shown to be effective in the context of taking eye images.

B. Trustworthiness

Our results indicate that providing users with disease-specific information increases the trustworthiness and potentially also positively impacts benevolence and integrity.

This is consistent with Akter et al. [42] who regarded the ability dimension, i.e. the perceived skills and competences of an app that encourage users to use it, as the most significant. Therefore, we confirm our hypothesis H1. On the other hand, trustworthiness is lowest when the assessment and the diagnosis modalities do not match, we can confirm H2. The disclaimer did not lead to the highest benevolence and integrity dimension, meaning we reject our hypotheses H3 and H4. There are descriptive trends that suggest that disease information is the most effective factor to increase these two trustworthiness dimensions. Furthermore, participants reported the highest trustworthiness when shown disease information; we therefore confirm H5.

C. Diagnostic Limitations

Whilst we propose that *EyeGuide* requires less expert knowledge and provides better accessibility compared to existing smartphone adaptors for retinal imaging, it is important to note that the spectrum of detectable diseases between the two approaches differs. Through fundus images taken using smartphone adaptors, a range of diseases such as diabetic retinopathy, glaucoma, and age-related macular degeneration can be diagnosed [47]; with these images, algorithmic approaches can be use to predict cardiovascular risk factors [48]. Smartphone cameras without additional lens adaptors do not allow for retina inspection and can only identify symptoms of diseases that are visible on the outside of the eye (e.g., conjunctiva or cornea). However, as some of the diseases visible with the naked eye, e.g. cataracts, are preventable [2], there is an obvious need for better screening, education and intervention in eye care. We argue that easy access to *EyeGuide* through a smartphone and the possibility of detecting cataracts, which are responsible for 33% of global blindness, with smartphone images justifies further research on this topic.

D. Contribution and Future Work

Given the recent development of AI-based diagnosis approaches, we argue that there are a large amount of potential application areas for *EyeGuide* to be integrated as part of existing eye-disease screening and diagnosis tools. The interactive tutorial, based on principles of multimedia learning, was successful in preparing most users to capture high quality images. Minor usability issues were identified in the user study and will be addressed in the following design iterations of *EyeGuide*.

We are one of the first to explore solutions for guiding image capture for eye images. While similar approaches in the domain of dermatology [30] include features to automatically trigger image capture, they do not provide any support for device positioning. Other authors [36] described the use of verbal guidance for taking portrait photos, however, our interactive tutorial (IT) approach combines a tutorial and verbal guidance as an integrated solution.

Based on our findings, we suggest that the designers of mHealth applications should include disease-specific information to ensure trustworthiness. While we focused on color

blindness in our trustworthiness study, our findings are generalizable to other mHealth fields beyond eye-related diseases. As future work we plan to extend our study to encompass a broader sample, e.g. the elderly, and to explore the use of our *EyeGuide* approach as part of iris imaging for biometric authentication [49, 50, 51].

VI. Conclusion

In this paper, we presented two studies exploring the usability and trustworthiness of eye-related mHealth applications. An interactive and multi-modal tutorial that demonstrates correct device positioning was successful in reducing acquisition time. An easy to use, self-administered app to support screening and diagnosis of eye-related conditions can increase eye care access, especially in developing countries. Providing disease-specific background information was shown to be the most effective intervention to increase trustworthiness in the diagnosis. Our findings are relevant to the designers and developers of mHealth applications in the area of ophthalmology and other medical fields.

References

[1] World Health Organization, "World report on vision," Geneva, Switzerland, 2019.

[2] World Health Organization, "Universal eye health: a global action plan 2014-2019," Geneva, Switzerland, 2013.

[3] World Bank Group, *World Development Report 2016: Digital Dividends*, ser. World Bank e-Library. Washington D.C., USA: World Bank Publications, 2016. [Online]. Available: https://books.google.de/books?id=dAl-CwAAQBAJ

[4] A. Russo, F. Morescalchi, C. Costagliola, L. Delcassi, and F. Semeraro, "Comparison of smartphone ophthalmoscopy with slit-lamp biomicroscopy for grading diabetic retinopathy," *American journal of ophthalmology*, vol. 159, no. 2, pp. 360–364, 2015.

[5] F. Lamonaca, G. Polimeni, K. Barbé, and D. Grimaldi, "Health parameters monitoring by smartphone for quality of life improvement," *Measurement*, vol. 73, pp. 82 – 94, 2015.

[6] N. U. Open Lab. (2018) Myskinselfie. [Online]. Available: http://www.myskinselfie.com

[7] M. Tomlinson, M. J. Rotheram-Borus, L. Swartz, and A. C. Tsai, "Scaling up mhealth: Where is the evidence?" *PLOS Medicine*, vol. 10, no. 2, pp. 1–5, 02 2013. [Online]. Available: https://doi.org/10.1371/journal.pmed.1001382

[8] S. R. Steinhubl, E. D. Muse, and E. J. Topol, "The emerging field of mobile health," *Science translational medicine*, vol. 7, no. 283, pp. 283–283, 2015.

[9] T. Greenhalgh, G. Robert, S. Bate, O. Kyriakidou, and F. MacFarlane, "How to spread good ideas: A systematic review of the literature on diffusion, spread and sustainability of innovations in health service delivery and organisation," 2004. [Online]. Available: http://www.sdo.nihr.ac.uk/files/project/38-final-report.pdf

[10] C. M. Angst and R. Agarwal, "Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion," *MIS quarterly*, vol. 33, no. 2, pp. 339–370, 2009.

[11] X. Guo, X. Zhang, and Y. Sun, "The privacy–personalization paradox in mhealth services acceptance of different age groups," *Electronic Commerce Research and Applications*, vol. 16, pp. 55 – 65, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1567422315000770

[12] F. Khatun, A. E. Heywood, P. K. Ray, S. Hanifi, A. Bhuiya, and S.-T. Liaw, "Determinants of readiness to adopt mhealth in a rural community of bangladesh," *International Journal of Medical Informatics*, vol. 84, no. 10, pp. 847 – 856, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1386505615300150

[13] K. Xu, D. Feng, and H. Mi, "Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image," *Molecules*, vol. 22, no. 12, p. 2054, 2017.

[14] H. Wei, A. Sehgal, and N. Kehtarnavaz, "A deep learning-based smartphone app for real-time detection of retinal abnormalities in fundus images," in *Real-Time Image Processing and Deep Learning 2019*, vol. 10996. International Society for Optics and Photonics, 2019, p. 1099602.

[15] C. Lam, C. Yu, L. Huang, and D. Rubin, "Retinal lesion detection with deep learning using image patches," *Investigative ophthalmology & visual science*, vol. 59, no. 1, pp. 590–596, 2018.

[16] S. K. Padhy, B. Takkar, R. Chawla, A. Kumar *et al.*, "Artificial intelligence in diabetic retinopathy: A natural step to the future," *Indian journal of ophthalmology*, vol. 67, no. 7, p. 1004, 2019.

[17] M. C. Roy, O. Dewit, and B. A. Aubert, "The impact of interface usability on trust in web retailers," *Internet research: Electronic networking applications and policy*, vol. 11, no. 5, pp. 388–398, 2001.

[18] F. N. Egger, ""trust me, i'm an online vendor": Towards a model of trust for e-commerce system design," in *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '00. New York, NY, USA: ACM, 2000, pp. 101–102. [Online]. Available: http://doi.acm.org/10.1145/633292.633352

[19] J. Kim and J. Y. Moon, "Designing towards emotional usability in customer interfaces—trustworthiness of cyber-banking system interfaces," *Interacting with computers*, vol. 10, no. 1, pp. 1–29, 1998.

[20] J. Rana and S. M. Galib, "Cataract detection using smartphone," in *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, Dec 2017, pp. 1–4.

[21] S. Pathak and B. Kumar, "A robust automated cataract detection algorithm using diagnostic opinion based parameter thresholding for telemedicine application," *Electronics*, vol. 5, no. 3, p. 57, 2016.

[22] Y. Alhasawi, B. Mullachery, and S. Chatterjee, "Design of a mobile-app for non-invasively detecting high blood cholesterol using eye images," in *51st Hawaii International Conference on System Sciences, HICSS 2018*, 2018, pp. 1–9. [Online]. Available: http://hdl.handle.net/10125/50295

[23] S. M. Kumar, R. Gunasundari, and N. Ezhilvathani, "Non-invasive measurement of cholesterol levels using eye image analysis," in *International Conference on Advances in Computational Intelligence and Communication (CIC)*, vol. 14, October 2016, pp. 33–42.

[24] A. Mariakakis, J. Baudin, E. Whitmire, V. Mehta, M. A. Banks, A. Law, L. Mcgrath, and S. N. Patel, "Pupilscreen: Using smartphones to assess traumatic brain injury," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, Sep. 2017. [Online]. Available: https://doi.org/10.1145/3131896

[25] F. Guo, Y. Mai, X. Zhao, X. Duan, Z. Fan, B. Zou, and B. Xie, "Yanbao: A mobile app using the measurement of clinical parameters for glaucoma screening," *IEEE Access*, vol. 6, pp. 77 414–77 428, 2018.

[26] M. Akil and Y. Elloumi, "Detection of retinal abnormalities using smartphone-captured fundus images: a survey," in *Real-Time Image Processing and Deep Learning 2019*, vol. 10996. International Society for Optics and Photonics, 2019, p. 109960K.

[27] T. N. Kim, F. Myers, C. Reber, P. Loury, P. Loumou, D. Webster, C. Echanique, P. Li, J. R. Davila, R. N. Maamari, N. A. Switz, J. Keenan, M. A. Woodward, Y. M. Paulus, T. Margolis, and D. A. Fletcher, "A Smartphone-Based Tool for Rapid, Portable, and Automated Wide-Field Retinal Imaging," *Translational Vision Science & Technology*, vol. 7, no. 5, pp. 21–21, 10 2018. [Online]. Available: https://doi.org/10.1167/tvst.7.5.21

[28] M. C. Munson, D. L. Plewman, K. M. Baumer, R. Henning, C. T. Zahler, A. T. Kietzman, A. A. Beard, S. Mukai, L. Diller, G. Hamerly, and B. F. Shaw, "Autonomous early detection of eye disease in childhood photographs," *Science Advances*, vol. 5, no. 10, 2019. [Online]. Available: https://advances.sciencemag.org/content/5/10/eaax6363

[29] D. Diethei and J. Schöning, "Using smartphones to take eye images for disease diagnosis in developing countries," in *Proceedings of the Second African Conference for Human Computer Interaction: Thriving Communities*, ser. AfriCHI '18. New York, NY, USA: ACM, 2018, pp. 34:1–34:3. [Online]. Available: http://doi.acm.org/10.1145/3283458.3283521

[30] Skinvision. (2018) Skin cancer melanoma detection app. [Online]. Available: https://www.skinvision.com

[31] D. Diethei, A. Colley, M. Kalving, T. Salmela, J. Häkkilä, and J. Schöning, "Medical selfies: Emotional impacts and practical challenges," in *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3379503.3403555

[32] S. Holland, D. R. Morse, and H. Gedenryd, "Audiogps: Spatial audio navigation with a minimal attention interface," *Personal and Ubiquitous computing*, vol. 6, no. 4, pp. 253–259, 2002.

[33] R. Absar and C. Guastavino, "Usability of non-speech sounds in user interfaces," in *Proceedings of the 14th International Conference on Auditory Display (ICAD '08)*, 06 2008.

[34] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and icons: Their structure and common design principles (abstract only)," *SIGCHI Bull.*, vol. 21, no. 1, pp. 123–124, Aug. 1989. [Online]. Available: http://doi.acm.org/10.1145/67880.1046599

[35] F. Heller and J. Schöning, "Navigatone: Seamlessly embedding navigation cues in mobile music listening," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–7.

[36] C.-C. Lo, S.-P. Huang, Y. C. Ren, and Y.-C. Tseng, "Yes, right there!: a self-portrait application with sensor-assisted guiding for smartphones." in *MobiSys*, 2013, pp. 505–506.

[37] J. Balata, Z. Mikovec, and L. Neoproud, "Blindcamera: Central and golden-ratio composition for blind photographers," in *Proceedings of the Mulitimedia, Interaction, Design and Innnovation*. ACM, 2015, p. 8.

[38] V. Vo, L. Auroy, and A. Sarradon-Eck, "Patients' perceptions of mhealth apps: Meta-ethnographic review of qualitative studies," *JMIR mHealth and uHealth*, vol. 7, no. 7, p. e13817, 2019.

[39] A. L. Fruhling and S. M. Lee, "The influence of user interface usability on rural consumers' trust of e-health services," *International journal of electronic healthcare*, vol. 2, no. 4, pp. 305–321, 2006.

[40] K. Oyibo, I. Adaji, and J. Vassileva, "What drives the perceived credibility of health apps: Classical or expressive aesthetics?" in *HealthRecSys@ RecSys*, 2018, pp. 30–35.

[41] I. A. Chatzipavlou, S. A. Christoforidou, and M. Vlachopoulou, "A recommended guideline for the development of mhealth apps," *mHealth*, vol. 2, pp. 21–21, May 2016, 28293597[pmid]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28293597

[42] S. Akter, J. D'Ambra, and P. Ray, "Trustworthiness in mhealth information services: an assessment of a hierarchical model with mediating and moderating effects using partial least squares (pls)," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 1, pp. 100–116, 2011.

[43] R. C. Clark and R. E. Mayer, *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons, 2016.

[44] OpenCV. (2018) Opencv: Face detection using haar cascades. [Online]. Available: https://docs.opencv.org/3.3.0/d7/d8b/tutorial_py_face_detection.html

[45] K. Karrer, C. Glaser, C. Clemens, and C. Bruder, "Technikaffinität erfassen–der fragebogen ta-eg," *Der Mensch im Mittelpunkt technischer Systeme*, vol. 8, pp. 196–201, 2009.

[46] S. Ishihara, *Test for colour-blindness*. Kanehara Tokyo, Japan, 1917.

[47] E. Roberts, R. Morgan, D. King, and L. Clerkin, "Funduscopy: a forgotten art?" *Postgraduate Medical Journal*, vol. 75, no. 883, pp. 282–284, 1999. [Online]. Available: https://pmj.bmj.com/content/75/883/282

[48] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. S. Corrado, L. Peng, and D. R. Webster, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nature Biomedical Engineering*, vol. 2, no. 3, p. 158, 2018.

[49] S. Thavalengal, P. Bigioi, and P. Corcoran, "Iris authentication in handheld devices - considerations for constraint-free acquisition," *IEEE Transactions on Consumer Electronics*, vol. 61, no. 2, pp. 245–253, May 2015.

[50] K. B. Raja, R. Raghavendra, C. Busch, and S. Mondal, "An empirical study of smartphone based iris recognition in visible spectrum," in *Proceedings of the 7th International Conference on Security of Information and Networks*, ser. SIN '14. New York, NY, USA: ACM, 2014, pp. 239:239–239:246. [Online]. Available: http://doi.acm.org/10.1145/2659651.2659704

[51] K. B. Raja, R. Raghavendra, M. Stokkenes, and C. Busch, "Multi-modal authentication system for smartphones using face, iris and periocular," in *Biometrics (ICB), 2015 International Conference on*. IEEE, 2015, pp. 143–150.