Understanding and Modeling Viewers' First Impressions with Images in Online Medical Crowdfunding Campaigns

Qingyu Guo
qguoag@connect.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

Siyuan Zhou
szhoubd@connect.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

Yifeng Wu
ywuec@connect.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

Zhenhui Peng
pengzhh29@mail.sysu.edu.cn
Sun Yat-Sen University
Zhuhai, China

Xiaojuan Ma
mxj@cse.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

ABSTRACT

Online medical crowdfunding campaigns (OMCCs) help patients seek financial support. First impressions (FIs) of an OMCC, including perceived empathy, credibility, justice, impact, and attractiveness, could affect viewers' donation decisions. Images play a crucial role in manifesting FIs, and it is beneficial for fundraisers to understand how viewers may judge their selected images for OMCCs beforehand. This work proposes a data-driven approach to assessing whether an OMCC image conveys appropriate FIs. We first crowdsource viewers' perception of OMCC images. Statistical analysis confirms that agreement on all five dimensions of FIs exists, and these FIs positively correlate with donation intention. We compute image content, color, texture, and composition features, then analyze the correlation between these visual features and FIs. We further predict FIs based on these features, and the best model achieves an overall F1-score of 0.727. Finally, we discuss how our insights could benefit fundraisers and possible ethical concerns.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**.

KEYWORDS

Online medical crowdfunding campaign, first impression, computational assessment

ACM Reference Format:
Qingyu Guo, Siyuan Zhou, Yifeng Wu, Zhenhui Peng, and Xiaojuan Ma. 2022. Understanding and Modeling Viewers' First Impressions with Images in Online Medical Crowdfunding Campaigns. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3491102.3501830

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3501830

1 INTRODUCTION

Online medical crowdfunding platforms (OMCPs), such as GoFundMe [42], Qingsongchou [95], and Shuidichou [108], offer an opportunity for patients and/or caregivers to seek financial support via the Internet [16]. Figure 1 is an example of an online medical crowdfunding campaign (OMCC) shared on Twitter. A patient or related persons, often the patient's family or friend(s), denoted as the fundraiser, can initiate an OMCC by selecting image(s) of patients and describing the reasons for seeking financial support. Then the fundraiser would share the OMCC on social media platforms to persuade viewers to further read the campaign in the OMCP in the hope of receiving donations.

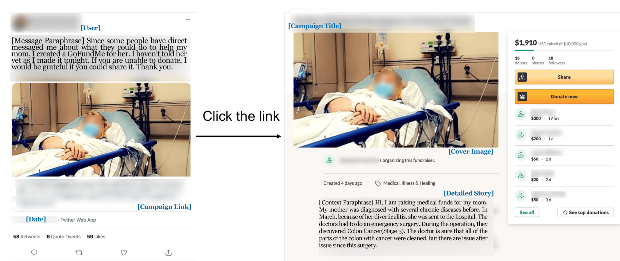


Figure 1: Example of a fundraiser sharing an OMCC on Twitter. The viewer can click the link to check the campaign's details in the OMCP and decide whether to donate. For privacy concerns, we have obscured sensitive information (apply to all images in this work). In this figure, we paraphrase the textual information to further protect user privacy.

Existing research indicates that first impressions conveyed by an online crowdfunding campaign could affect viewers' donation decisions to a great extent [24, 105]. From viewers' perspective, the perceived first impressions determine their initial interest in a crowdfunding campaign, and proper first impressions would encourage them to explore more facts about the campaign (e.g., clicking the link as shown in Figure 1) [135]. Whether or not viewers decide to make a contribution may depend on the details they read, but bad first impressions are likely to lose their attention in the first place [115]. Proper first impressions of OMCCs, as suggested by [23, 37, 69, 74, 114], could stimulate viewers' feelings of empathy, credibility, justice, impact, and attractiveness. For example, making viewers feel empathetic towards a medical campaign is positively correlated with their donation intentions [74].

To ensure the success of OMCCs, it is essential to offer assistance to fundraisers to improve the delivery of proper first impressions. Images play a critical role in building the first impression for viewers [24, 58] and could influence their subsequent behaviors (e.g., whether to read the detailed description of the campaign) and ultimately the monetary contribution. Traditionally, fundraisers can receive general guidance from OMCPs about preparing the images (especially the cover image) for the OMCC to present the patients' conditions and needs to the viewers. Take the GoFundMe platform as an example, the platform prompts users that photos should be visually appealing and high-quality [43]. However, fundraisers often have to anticipate viewers' perception towards their choice of photos based on their limited experience [61, 63, 88], and the impressions they try to manifest may not match the expectations of potential donors, e.g., they may overemphasize the emotional narratives for viewers [62]. Although previous works have indicated that certain image elements, such as the patient's gender [98] and facial expression [131], are correlated to the perceived first impressions and could influence the performance of OMCCs, it is still difficult for fundraisers to implement such empirical suggestions. On the one hand, some elements (including the gender of the patient) can not be changed by fundraisers. On the other hand, the effect of one image on perceived impressions is usually the interplay of various visual features [60, 124]. In brief, it would be beneficial for the OMCC fundraisers to understand what contents could be shown in the images and how viewers might judge their campaigns [63], but few practical guidelines are readily available.

The success of computational image assessment in other scenarios concerning user perception, such as brand personality [127] and animation engagement [126], motivates us to explore the possibility of deriving a computational model of general viewers' first impressions of medical campaign images. To this end, we first sample 450 OMCC images from a large collection of campaign cover pictures on the GoFundMe website. We then gather crowdsourced user perception ratings on different dimensions of first impressions (i.e., empathy, credibility, justice, impact, and attractiveness), and viewers' donation intent given the OMCC images. With this annotated image dataset, we seek answers to the following research questions: **RQ1)** Can viewers form consistent first impressions and projected donation intentions given OMCC images? **RQ2)** What are the relations between different first impressions and the perceived intent to donate? **RQ3)** What are the relations between visual features and different first impressions? **RQ4)** Can we build a computational model for predicting viewer's first impressions of OMCC images? And if yes to RQ4, **RQ5)** What are the visual features that contribute the most to predicting the first impressions of OMCC images?

Our analysis of ratings obtained from crowd workers shows that consistent impressions towards an individual OMCC image could exist. Moreover, these impression ratings are positively correlated with the donation intention, which implies the possibility of evaluating candidate pictures on these five dimensions of the first impressions to help the fundraiser select a proper cover image for an OMCC. Next, we collect a series of image features, including content-based, color-based, texture-based, and composition-based features, and examine their relations with first impression ratings by Spearman correlation analysis. Results suggest that all four feature sets are correlated with viewers' first impressions toward OMCC images, which is also confirmed by crowd workers' rationale for their ratings. We then formulate the first impression modeling as a classification task (above/below mean of crowdsourced ratings in each impression), and train multiple models. The best one, Random Forest, achieves an overall F1-score of 0.727. By combining results from model feature importance analysis and the previous correlation analysis, our work highlights the essential role of image content features and color features for the manifestation of first impressions. We acknowledge that our findings might be exploited for deception in OMCCs. As indicated in the literature, ethical concerns exist in the context of OMCCs, including fraudulent and abusive use of the campaign and privacy disclosure [110, 111, 133]. However, very little work has discussed the possible ethical impact of the image element – the focus of this paper – in a medical campaign. We summarize the ethical concerns of our work, elaborate on possible solutions to address them, and discuss design implications for fundraisers and OMCPs.

Our main contributions are three folds:

- We conduct crowdsourced experiments to investigate how humans perceive images in medical campaigns and explore computational models of viewers' perceptions using data-driven methods.
- We compile a list of first impressions related to donation behavior in OMCCs based on the literature survey and analyze the relations between these first impressions and the donation intention.
- We analyze the relations between visual features and viewers' first impressions, and model viewers' first impressions for OMCC images via the data-driven approach. We also discuss the ethical implications of our findings. Our insights could benefit OMCC fundraisers in better preparing campaigns and have implications for the future design of OMCPs.

2 RELATED WORK

Crowdfunding is a practice of raising funds for a campaign from a large number of people, typically through the Internet [9]. Online crowdfunding campaigns can be classified into different categories by their purposes. About one-third of campaigns on a typical crowdfunding platform are financing for medical expenses [79]; they are denoted as online medical crowdfunding campaigns (OMCCs) in this paper. Compared with other types of crowdfunding such as those soliciting funds for a creative project or a start-up business, OMCCs are donation-based and donors have no expectation of reward [107]. Common examples of online medical crowdfunding platforms (OMCPs) are GoFundme, Shuidichou, and Qingsonghcou. GoFundMe has the greatest impact in scale [98].

In this section, we first introduce the role of first impressions for OMCCs and survey the core first impressions related to donation behaviors. This forms the research foundation of this paper. Then

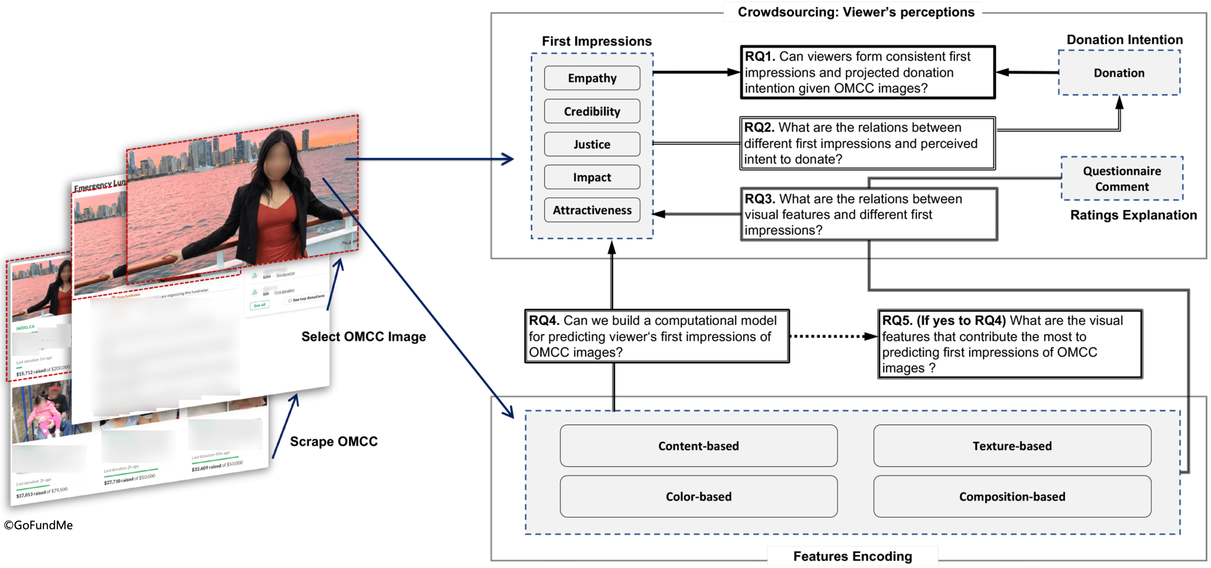


Figure 2: The workflow of this study. We first scrape OMCCs from the GoFundMe platform and sample OMCC images. We then crowdsource viewers' first impressions and donation intention toward the given image. We analyze the received ratings, and the correlation between viewers' perceived first impressions and donation intention. We encode visual features, and analyze the relations between these visual features and the perceived first impressions. Finally, we model the first impressions of OMCC images with the extracted and manually annotated features, and analyze contributing factors for prediction.

we demonstrate the importance of image for first impression manifestation in OMCCs, and identify the existing gaps for fundraisers to prepare the image for OMCCs. Finally, we investigate the existing applications and techniques of visual perception modeling.

2.1 First Impressions of OMCC

First impressions are vital for decision-making in multiple domains of human-computer interaction, such as commercial crowdfunding [135] and online tourism [76]. Establishing the proper first impressions about the crowdfunding project is the preliminary step for viewers to explore further details about the campaign, and eventually, they might consider making monetary contributions [135]. As there is no existing framework on what are critical first impressions for OMCCs, in this paper, we borrow theories that may influence viewers' donation behaviors in the domain of pro-social behaviors, charity behaviors, and donation-based crowdfundings. According to our survey, first impressions that might be correlated with viewers' donation intentions include perceived empathy, credibility, justice, impact, and attractiveness [23, 37, 69, 74, 114]. This section introduces the meaning of these five first impressions and their potential relationship with the donation intention.

The perceived empathy measures the level of compassion felt by the potential donor for the campaign owner or patient [52]. Past literature in pro-social behavior has theorized perceived empathy from the emotional state (i.e., feel what others feel) and cognitive perspective (i.e., understand and evaluate others' situation and intention) [26, 27, 35, 38, 53]. In accordance with both definitions, past research suggests that perceived empathy can prompt people's pro-social behavior (e.g., involving in charitable causes) [6, 7, 23, 35, 69]. In the scenario of OMCCs,different components of a campaign are associated with perceived empathy, such as campaign's popularity [74] and facial emotions of people in the picture [131].

The perceived credibility for OMCCs refers to potential donors' trust towards an online campaign, which improves donation results through enhancing persuasion [123]. As suggested by literature in the domain of charity, potential donors tend to trust the campaigns owned or endorsed by their friends [63]. Kim et al. [62] further identify the picture as an important source for viewers to establish credibility towards the OMCC. Therefore, we investigate whether and how the perceived credibility could be formed solely based on the OMCC image.

The perceived justice reflects whether the potential donors understand the fundraisers' situation, and further feel it is fair to donate to them [69]. The perceived justice would positively influence the donation intention if the fundraiser successfully makes potential donors feel deserving to help them and vice versa [131]. For example, viewers may feel it is unjust to donate to the fundraiser if viewers think the fundraiser's fault causes the current status.

The perceived impact, also known as perceived utility or efficacy, is a mechanism used in charitable giving, which stands for whether viewers feel that help could make great differences for the target campaign [31, 37]. Campaigns that require a greater amount of help could make viewers feel less sense of impact as the immediate effect would not take place after the donation, and may demotivate the donation intention. Such a phenomenon can be explained by the theory of "collapse of compassion", which means people tend to be less enthusiastic about helping when they feel costly support is expected, e.g., feeling compassionate for one person in pain rather than worrying about eight people in suffering [17].

The perceived attractiveness refers to how viewers perceive the visual appeal that arouses interest in an object [71]. For example, Choi et al. [21] show how the background color connects with viewers' attractiveness towards the charity campaign and its impact on charity performance. In the context of OMCCs, perceived attractiveness affects potential donors' intention to read the campaign, which is the prerequisite for a donation decision to make.

These dimensions of first impressions could be correlated. For example, it is suggested that the perceived attractiveness is positively correlated with the perceived credibility of a website [100]. However, the level of different first impressions might be inconsistent toward the same campaign, e.g., the contradiction could exist in the dimension of empathy and justice [69]. Nevertheless, few works systematically analyze the relationship between these first impressions and the donation intention. In this work, we adopt the standardized measurement of each type of impression and investigate the mutual effect on donation intentions from these first impressions. In addition, previous works generally take the multimodality information in campaigns, including both textual and visual features. However, it is not clear whether visual features could construct impressions to viewers in medical campaigns. This study examines the role of images in conveying impressions in the context of OMCCs.

2.2 Research of image for OMCC

Previous studies have suggested that content and sharing on social networks are correlated with the campaign performance. It has been figured out that the social tie between a donor and a fundraiser [28], the visual cues (e.g., images) [98, 99], and the textual cues (e.g., title, descriptions) [134] would influence the performance of OMCCs, among which fundraisers have greater control over the latter two factors. Compared with textual cues, images may attract viewer's attention easier, expedite information processing, and be manipulable to shape certain attitudes [3, 58, 82]. The selection of images has been investigated in the domain of advertising, marketing, e-commerce [5, 20, 56], and it is suggested that a proper image is the key to persuading viewers [60].

Past research examines the statistical correlation between visual elements of OMCC images (such as the image theme, the facial expression, age, gender, and the number of people) and the campaign performance [68, 98, 99, 123, 131, 138]. For example, it has been indicated that the number of people in an image is positively correlated with the percentage of donation goals achieved [138], while the average percentage of goals achieved is less for campaigns targeting males than females [98]. Some works further investigate the role of the cover image in shaping viewers' first impressions and affecting the performance of an OMCC. For instance, Wang et al. [123] indicate that photographic narratives on the patient's healthiness or unhealthiness can evoke positive emotions (e.g., hope, empathy) of viewers, which might encourage the donation. Yazdani et al. [131] suggest the positive correlation between the patient's expressed emotions from the photography and the viewers' donation intention, where viewers' perceived empathy and justice can mediate the correlation.

Nevertheless, it has been suggested that fundraisers might have difficulty in selecting images for the medical campaign to manifest their expected first impressions [61, 63]. For instance, prior research finds it challenging to convey the justice and credibility of a campaign through images when the medical condition is not visually noticeable (e.g., losing the eyesight) [61]. Although the correlation between some visual features and certain types of first impressions (e.g., perceived empathy, credibility) as well as the fundraising performance have been identified in the literature, these empirical findings are often impractical for fundraisers to follow. First, some findings are about attributes of the patient (e.g., gender, age), which can not be changed by the fundraiser. Second, the perceived first impressions of an image are usually a combined effect of various visual elements [60, 124]. Very little research has comprehensively examined the efficacy of various kinds of image features in impression delivery under the context of OMCCs. In short, previous works leave little opportunity for fundraisers to receive explicit feedback on the image they would like to choose. Instead, our work focuses on computationally modeling the perceptions received by the viewers given an image of a medical campaign, which makes it possible for instant evaluation of OMCC images.

2.3 Computational modeling of Visual Impression

There have been attempts to model viewers' perceptions of visual stimuli in multiple scenarios, such as the colorfulness and usability of the website [89, 97], the brand personality and visual engagement of mobile UI [127, 128], and the aesthetic of infographics [48]. For example, Reinecke et al. [97] model perceived aesthetics of the website through the website's colorfulness, visual complexity, and user's demographic information with a linear mixed-effects model. Wu et al. [126] model user engagement given UI animations through a deep learning framework with temporal and spatial features as input. Meanwhile, previous works show the possibility of modeling perceptions toward human pictures. For instance, Vernon et al. [119] illustrate that first impressions of human social traits (e.g., perceived attractiveness) can be predicted from the photography of faces by feeding facial attributes into neural networks. Joo et al. [60] model the perceived emotion and personality of politicians appeared in the mainstream media through rank SVM with the input of facial display, gestures, and the scene context. In general, previous studies model the perceptions either with visual features through machine learning models, or directly train deep learning models with the image pixels as input in an end-to-end fashion. Some machine learning approaches (e.g., tree-based methods) are often more interpretable than deep learning methods, while the latter might outperform the former on large datasets [44, 136]. This research mainly focuses on modeling first impressions with machine learning models to derive more practical implications.

Although the success of previous works in visual impression modeling shed light on the prediction of first impressions for OMCC images, the aspects of feature sets adopted in past works are usually relatively small, which may limit the ability to depict human images and scale of findings. Moreover, compared with previous studies in modeling perceptions for human images, the first impressions of OMCC images include unique dimensions, such as the perceived empathy and justice. Therefore, the relationships between visual features and first impressions, and whether first impressions are predictable with these visual features have not been fully examined.

To address the gaps, we derive and customize comprehensive visual feature sets to describe OMCC images, and investigate how first impressions are correlated with visual features. We further explore to model the first impressions with the visual features through machine learning models.

Summary of research gaps: According to the theories in the OMCC-related domains (e.g., pro-social behaviors), we summarize five first impressions (i.e., perceived empathy, credibility, justice, impact, and attractiveness) that might correlate with viewers' donation intentions in OMCCs [23, 37, 69, 74, 114]. However, it is unclear whether viewers can form consistent first impressions solely based on the OMCC images and the relationships between different first impressions and donation intentions. We address the former questions via crowdsourcing in **RQ1**, and examine the relationships through statistical analysis in **RQ2**. Although obtaining a set of visual features to depict images for impression modeling is essential, it is under-explored in OMCC images. Moreover, the relationships between visual features and five first impressions have not been fully investigated. In addition, it is unknown to what extent first impressions for OMCC images can be predicted with these features. To fill the gaps, we first summarize and customize visual features in related domains, including content-based features, color-based features, texture-based features, and composition-based features. Then we explore the correlation between these visual features and first impressions perceived by viewers (**RQ3**). Next, we build machine learning models with derived visual features to explore the predictability of first impressions of OMCC images (**RQ4**), and further identify contributing visual factors in modeling first impressions (**RQ5**). This work contributes to a deeper understanding of the role of first impressions conveyed from OMCC images, and presents how to model the first impressions.

3 DATASET COLLECTION AND PERCEPTION ASSESSMENT

We introduce the research site we investigate and how we create the OMCC dataset. We then introduce details of the survey questionnaire as well as crowdsourcing procedures.

Ethics and Privacy Protection: Prior to this work, our research team obtained IRB approval for data collection and analysis. We adopt several protocols to protect the privacy of people who are included in the campaign image. First, we disable the download of the images by preventing right click on the MTurk website. Second, each image can only be displayed once for a limited short time to ensure participants will not have enough time to take screenshots. Third, we remove such data from the MTurk database soon after the closure of the study to eliminate the data leakage issue.

3.1 Research Site and Dataset

We use data from the GoFundMe platform, one of the largest OMCPs. In order to retrieve diverse medical campaigns in the US, we first identify a list of disease-related keywords, then search for related campaigns following the procedure in [66]. Specifically, we refer to the common critical illnesses with heavy expenses in the US, including cancer, heart attack, and stroke, which are the leading causes of death in the US [1] and are within the coverage of common US critical illness insurance [51]. Then, we collect common subcategories of these illnesses from the government websites [19, 59] and medical-related literature [120]. Finally, we obtain 29 keywords of critical illnesses which could have the necessity of raising crowdfunding, and the complete list of keywords is shown in Table 1.

Table 1: Query words for scraping campaigns from the GoFundMe platform.

Disease category	Subcategory query word
1- Cancer	*Breast Cancer, Lung Cancer, bronchus cancer, prostate cancer, colon cancer, rectal cancer, bladder cancer, non-hodgkin lymphoma, kidney cancer, renal pelvis cancer, endometrial cancer, leukemia, pancreatic cancer, thyroid cancer, liver cancer*
2- Heart Disease	*congenital heart disease, kawasaki, arrhythmias, Sudden Cardiac Arrest, Atherosclerosis, Coronary Heart Disease, Cardiomyopathy, Heart Failure, Valvular Diseases, Peripheral Artery Disease, microvascular, Aortic Diseases*
3- Stroke	*Ischemic stroke, Hemorrhagic stroke*

We then scrape campaign URLs and associate information with Python library Beautiful Soup in June 2021. The GoFundMe website supports requesting a maximum of $1,000$ campaigns for each query word, and we successfully collect $11,575$ unique campaigns by querying with these 29 query keywords. We collect the cover image, the tag, the location of the fundraiser, and the created date associated with each campaign. Each OMCC contains only one 720×405 pixels cover image in JPG format. We exclude campaigns not located in the US or created before the year of 2017 to minimize the spatial and temporal differences. We further remove campaigns that are not with the tag of "*Medical, Illness & Healing*", whose main purposes are not seeking medical help, such as appealing for education donation, whereas containing cancer in the content. After this step, we obtain a dataset of $7,039$ OMCCs for critical illnesses in the US.

3.2 Theme Distribution of OMCC Images

An implicit question before sampling images for annotation is how fundraisers use images for medical crowdfunding on the GoFundMe platform. Three researchers familiar with the GoFundMe platform conduct a thematic analysis for images in the OMCC dataset. Two of them first randomly sample 100 images from the entire OMCC dataset, then group them into several clusters and name the theme for each cluster independently. Next, they meet regularly to compare and discuss images in each cluster, and refine the cluster theme. After several rounds of iterative clustering and discussion, they agree with the coding scheme: *healthiness narratives with a single person*, *healthiness narratives with multiple people*, *unhealthiness narratives*, *collage*, and *not contain human*. Examples of themes are illustrated in Figure 3.

To understand the theme distribution of the entire dataset, another 800 images are randomly sampled from the entire dataset for training an image category classifier following [123]. These two researchers process the image classification independently according to the discussed coding themes. They also assign a theme *others* if they feel the image does not fall into any theme mentioned

(a) Healthiness narrative with a single person (b) Healthiness narrative with multiple people (c) Unhealthiness narrative (d) Collage (e) Non Human

Figure 3: Example image in each theme. Images © GoFundMe.

Table 2: Distribution of our medical campaign image databse.

Theme	(a) healthiness narrative with a single person	(b) healthiness narrative with multiple people	(c) unhealthiness narrative	(d) collage	(e) non human
Number	3123	2010	1426	244	108
Percent	45.2%	29.1%	20.6%	3.5%	1.6%

above. The inter-rater metric Cohen's κ is 0.88, indicating a strong agreement [80]. The third researcher screens the images with disagreement and decides the final label for those images. We split these images into a training set (60%), a validation set (20%), and a test set (20%), respectively. A multi-class classifier is proposed with the standard ResNet-34 network [50] (pre-trained on the ImageNet dataset [29]) as the backbone, followed by a fully connected layer. The classifier is implemented based on the Pytorch [93], and achieves an accuracy of 77.9% on the validation set and 78.2% on the test set. We save the model and predict the image category for the entire database. The detailed distribution of images in each theme is listed in Table 2.

3.3 Crowdsourcing Procedures

3.3.1 Image Selection. We first manually filter out the misclassified images in the database. Then we randomly sample 120 images from each of the three most commonly used image themes (Table 2(a, b, c)), and randomly sample 45 images from each of the two least frequently used image themes (Table 2(d, e)) for diversity. In the end, we select 450 images for annotation.

3.3.2 Survey Questionnaire. We adopt standardized measurement of each dimension of first impressions (i.e., perceived empathy, credibility, justice, impact, and attractiveness) and the donation intention from literature in related domains, such as online crowdfunding and charity donation. We tailor these questions according to our setting. Each question starts with "The presented image makes you feel ...", and is measured on a 7-point Likert scale, ranging from "1 - Strongly Disagree" to "7 - Strongly Agree". We list the questions in Table 3. In total, there are 19 questions.

3.3.3 Crowd Ratings. To collect labels for the perceived first impressions and donation intentions given the image, we develop crowdsourcing tasks on Amazon Mechanical Turk. We restrict crowd workers to be located in the US to eliminate cultural differences across countries. The participants first report their basic demographic information, including age and gender. Next, three images are randomly assigned to each participant for evaluation one by one. We follow the strategy of five-second tests in the UX research [33] for each task to measure the first impressions. In detail, each image is displayed for five seconds, then the image disappears,

Table 3: Detailed description of our survey questionnaire.

Dimensions	Items	Source
Perceived Empathy	Sympathetic Warm Compassionate Soft-hearted Tender Moved	[69, 73]
Perceived Credibility	The campaign is trustworthy The campaign is reliable The campaign is credible	[91]
Perceived Impact	The donation can do a lot of good The donation can make a big difference The donation is positive with expected cosequences	[37]
Perceived Justice	Donating to this campaign is fair Donating to this campaign is just	[69, 131]
Perceived Attractiveness	Will pay much attention to this campaign The campaign is attention grabbing The campaign is effective	[21]
Donation Intention	I would feel better if I donate My intention to donate to this campaign is high	[32, 137]

and the survey questionnaire for measuring impressions and intentions (Table 3) is presented. The participants are not allowed to view the image for the second time so that they can report based on their first impressions.

In each task, participants need to answer two identical questions on perceived impressions and are required to select a specific option for quality control. The participants are further asked to explain their ratings for the given image. We filter out invalid tasks if the answers meet any of the following criteria: 1) fail to pass the quality control questions 2) consistent patterns in ratings or explanations 3) meaningless explanations for ratings 4) unreasonable short completion time (less than 50 seconds). Participants will be given 0.14 USD as a reward for every valid task (maximum three tasks), and the average duration for each task is about two minutes. After removing invalid tasks, we sort responses for each image in chronological order, and only keep the first five responses. For images that remain less than five valid responses, we repeat the above-mentioned crowdsourcing steps until the requirement of five valid responses for each image has been satisfied. Eventually, each image is labeled by five different participants. In total, we recruit 1,227 participants (55.2% female, age mean ± SD = 37.6 ± 12.3, age range 18 − 77). The largest age group of participants (41.9%) falls under the age of 25 − 34 years old, followed by the age range of 35 − 44 (22.6%). We average the ratings from five participants for each image to reduce the possible effect of individual rating preference. The average standard deviation of five rating scores toward each image is in the range of 0.999 − 1.211 for all measured dimensions. Details of the rating statistics are reported in Section 5.1.

4 METHODS

Our work aims to model viewers' first impressions given OMCC images. In this section, we first introduce the adopted visual features. Then, based on the extracted and manually annotated visual features, we demonstrate how we define the first impression modeling task and the computationally modeling approaches.

4.1 Predictive Visual Features

Image low-level features have been extensively used for forecasting perceived high-level impressions, such as website aesthetics, mobile UI engagement [97, 126]. However, few works model the dimensions of impressions discussed in this paper. We first survey commonly used features for impression management and user perceptions that may be correlated with our investigated first impressions, then organize and list these features in Table 4.

4.1.1 Content-based features. Previous works suggest the content of the image could affect viewers' first impressions [62, 112, 124] and the donation performance [98]. The content-based features include three types: the human-related features (description for human in the image), the location-related features, and the theme-related features (high-level description of the image). The human-related features and image theme features are automatically extracted, while the location-related features are manually annotated[1]. In detail, for human-related features, we consider the human number, age-related features, gender-related features, emotion-related features, and facial attributes, which have shown correlation with perceived impressions and/or donation performance by literature [68, 98, 112, 122, 124, 131, 138]. We compute these human-related features with the Amazon Rekognition API [106], a widely adopted face detection and analysis API recommended by [130]. The Amazon Rekogition API can analyze face attributes for each detected human in an image with reliable accuracy. After extracting these human attributes, we calculate the above-mentioned features. As for location-based features, we focus on whether the image is captured indoor or outdoor, and for those indoor images, whether it presents hospital scenery. These two binary features are manually annotated. Lastly, the theme-related features refer to the five themes we discussed in Section 3.2.

4.1.2 Color-based features. Previous works have found color influences human emotions [77]. Various color-related features are examined for their effectiveness in inferring people's impressions. For example, the hue in photographs has an impact on the popularity of Instagram posts [132], while both saturation and brightness affect the invoked emotion [87]. The semantic color area distribution refers to the area of 11 basic colors in the image by learning from real-world objects [116]. Similarly, the HSV area distribution measures the area of pixels that fall into a different level of brightness, saturation, and hue [125]. These two color distribution features provide different aspects to describe the image color, which are both used for emotion classification [77].

4.1.3 Texture-based features. Previous studies show that texture can influence people's impression formation on faces and can help evaluate the attractiveness of images [102]. Tamura texture is suited to capture the emotional perception of visual textures [72]. We use the feature of coarseness, contrast, and directionality, which can capture the high-level perceptual attributes of a texture and are widely used in visual art appreciation [72]. Wavelet features are used to examine face attractiveness in the previous study [113]. GLCM features are commonly used for texture analysis [47]. For each image, we calculate contrast, correlation, energy, and homogeneity by the GLCM function.

4.1.4 Composition-based features. The Previous study has found the link between image composition and image aesthetics [84]. Level-of-detail can be used to assess image quality [34], which influences the performance of OMCCs greatly [138]. A low depth of field is the small or narrow area in an image that is in focus, which is used in professional photographs for shooting single objects by using larger aperture settings, macro lenses, or telephoto lenses. Dynamics describes the degree of motion for objects in the picture. We follow the computation in [77] to classify dynamic and static lines. Rule of thirds states that the main body of the photograph should be positioned in one of four intersections of four lines that segment the image into nine equal rectangles. Rule of thirds is also used for modeling image emotion [55].

4.2 Computational Models

Based on the extracted and manually annotated visual features, we propose a series of computational models for assessing the extent to which an image can manifest first impressions to a viewer. Since assessing whether the selected image for an OMCC conveys proper first impressions could benefit fundraisers [62, 63], we formulate the first impression prediction task as a classification task. For each dimension, images with the mean of five received ratings greater than the mean of overall ratings are labeled as "above average" on that specific dimension, and those below are labeled as "below average". For example, an "above average" image in the dimension of empathy represents that the conveyed empathy from that image is above the average level of empathy of our annotated images.

We adopt five commonly-used machine learning algorithms, including K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Multi-Layer Perceptron classifier (MLP), Decision Tree, and Random Forest for classification. We select these five machine learning algorithms for two reasons. First, previous works have shown the power of these algorithms in modeling first impressions given the visual stimuli in different scenarios [60, 78, 101, 119, 127], such

[1] In the future, the manually annotated features can be computed automatically with training data and the deep learning technique.

Table 4: Adopted visual features.

Feature Set	Feature Name	Number	Description	Source
Content-based	Human number	1	the number of people in the photo	[98, 138]
	Age-related	3	the minimum, mean, and maximum age of people in the photo	[98, 138]
	Gender-related	3	the number of female and male, the ratio of female in the photo	[98, 138]
	Emotion-related	6	the mean of six basic emotion for people in the photo, including happy, sad, angry, disgusted, surprised, fear	[131]
	Facial-related	5	the ratio of people that are smiling, opening the eye or mouth, wearing sunglasses, or normal glasses	[112, 124]
	Location-related	2	whether is captured indoor/outdoor, (not) in hospital	[112]
	Theme-related	5	five themes proposed in Section 3.2	
Color-based	HSV statistics	6	mean of saturation and brightness, hue distribution in 4 dimensions	[77]
	Semantic color area distribution	11	percentage of black, blue, brown, green, gray, orange, pink, purple,red, white, yellow pixels	[116]
	HSV area distribution	10	based on Wang features, area of brightness (very low, low, middle, high, very high), area of saturation (high, middle, low), area of hue (warm, cool)	[125]
Texture-based	Tamura	3	three tamura features (coarseness, contrast, directionality)	[102]
	Wavelet	12	wavelet textures(spatial smoothness/ graininess) in three levels on each HSV channel, sum of all levels in each channel	[113]
	GLCM	12	contrast, correlation, energy, homogeneity for three HSV channels	[47]
Composition-based	Level of Detail	1	segments number after waterfall segmentation	[34]
	Low Depth of Field	3	ratio of wavelet coefficients of inner rectangle to the hue, saturation, brightness	[25]
	Dynamics	6	the line slope of static, dynamic (absolute and relative), lengths of static lines, and lengths of dynamic lines	[77]
	Rule of thirds	3	mean HSV value of image's inner rectangle	[55]

as predicting human social attributes from photos and brand personality from mobile application screenshots. Second, they are representative and cover a wide range of classification algorithms. Specifically, KNN is a distance-based model, while SVC and MLP use non-linear kernels. Decision Tree is a tree-based model, and Random Forest is an ensemble of decision trees with the bagging method. These machine learning models take the extracted and manually annotated visual features listed in Table 4 as input, and are implemented with scikit-learn packages [94]. In addition, we propose a deep learning-based model for comparing the effectiveness of extracted and manually annotated visual features. The deep learning-based model adopts the ResNet-34 as the backbone, which is pre-trained on the ImageNet dataset, followed by a fully connected network. We implement the ResNet-34 based model with Pytorch [93]. For the deep learning-based model, we first resize the image to 224×224 pixels, then feed the pixels into the deep network.

To quantify the performance of classifiers, we adopt standard metrics of precision, recall, and F1-score. We randomly split these images into a training set (80%) and a testing set (20%). We determine the hyper-parameters of machine learning models by 10-fold cross-validation on the training set. In detail, the nearest neighbor number is set as 7 for KNN, the SVC adopts a sigmoid kernel, and the MLP has a one-layer hidden layer with RELU as the activation function. The depth for the decision tree is set to 6, while the RF ensembles 35 trees, whose depths are limited to 6. As for the ResNet-34 based model, we first split 25% of the training data as the validation set, and use the early stopping on the validation set for selecting the best hyper-parameters [18]. The deep learning-based model utilizes a two-layer fully-connected neural network (20, 2 perceptrons in each layer) with RELU as the activation function. After extracting the embedding from the ResNet-34 backbone, we randomly drop out 50% of neurons to alleviate the overfitting issue. The overall performance of each classifier, i.e., mean and standard deviation of classification metrics across five dimensions of first impressions on the test set, are reported in Section 5.4.

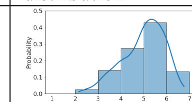
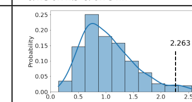
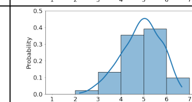
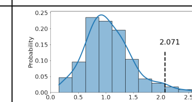
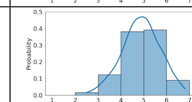
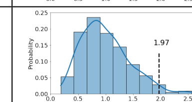
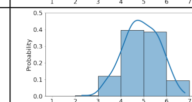
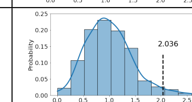
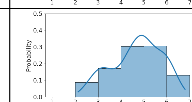
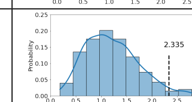
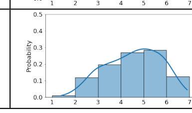
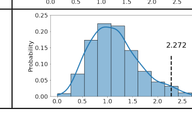
5 ANALYSIS AND RESULTS

In this section, we first verify whether viewers can form consistent first impressions and donation intention solely based on OMCC images. Then, we examine the correlation between first impressions and donation intentions. We further convey insights on what and how visual features may influence first impressions by analyzing the Spearman correlation between features discussed in Section 4.1 and perceived first impressions, and referring to crowd workers' explanations for their ratings. Finally, we attempt to model first impressions computationally, and identify contributing factors for our models.

5.1 Statistical Analysis for Ratings

In this section, we aim to understand to what extent participants could form consistent first impressions after observing an OMCC image for five seconds (**RQ1**). As there are multiple questions for the same item, we calculate the Cronbach's alpha value to measure the internal consistency of ratings for each item [13]. The Cronbach's alpha values for all items are greater than 0.85 as listed in Table 5, indicating very reliable ratings for perceived first impressions and donation intentions. We thus calculate the average of ratings for each item of a task. For each measured dimension, the average score of the overall ratings ($N = 2,250$) is around 4.8, and the

Table 5: Statistic information of ratings. The overall rating stands for the maintained valid ratings ($N = 2,250$), and the per image SD stands for the standard deviation of the five scores for each of the annotated images ($N = 450$). The dashed line in the column of "Per image SD distribution" stands for the top 2.5% standard deviation of the corresponding dimension.

	Cronbach's alpha	Overall rating		Per image SD		ICC	Overall rating distribution	Per image SD distribution
		Mean	SD	Mean	SD			
Empathy	0.945	4.990	1.376	1.038	0.476	0.694		
Credibility	0.953	4.892	1.330	1.053	0.431	0.648		
Impact	0.911	4.884	1.275	0.995	0.434	0.655		
Justice	0.874	4.893	1.251	0.999	0.444	0.608		
Attractiveness	0.92	4.700	1.591	1.211	0.518	0.697		
Donation	0.909	4.538	1.632	1.145	0.489	0.783		

average standard deviation of the five scores for each of the images ($N = 450$) is around 1. Table 5 depicts details of descriptive statistics and distribution of overall ratings as well as the standard deviation of received scores for each of the images.

We further evaluate the cross-user consistency on each measured dimension with the Intra-Class Correlation (ICC). In our experiment, each image is rated by $k = 5$ randomly assigned participants. Therefore, we calculate the ICC(1, k), which measures the absolute agreement between k raters for each question following the guideline in [67]. According to the standard criteria [22], the crowdsourcing participants reach good agreement on all five first impressions (ICC(1, k) > 0.6) and excellent agreement on donation intention (ICC(1, k) > 0.75) solely based on the medical campaign images. Among the five impression items, the perceived empathy and attractiveness achieve higher agreement, possibly because they are formed primarily through the affective route [21, 27]. The donation intention rating gains the highest level of consistency across viewers, indicating the importance of choosing a proper cover image for an OMCC, which is in line with the findings of previous works [123, 131]. However, we observe ICC(1,k) of images whose standard deviation of the received five ratings exceed 97.5% of images is smaller than 0.2 on any dimension of first impressions, suggesting poor agreement across raters for these images. This indicates although annotators generally have consistent ratings of first impressions towards OMCC images, controversial ratings might exist for certain images. We will elaborate on this point in Section 5.4.

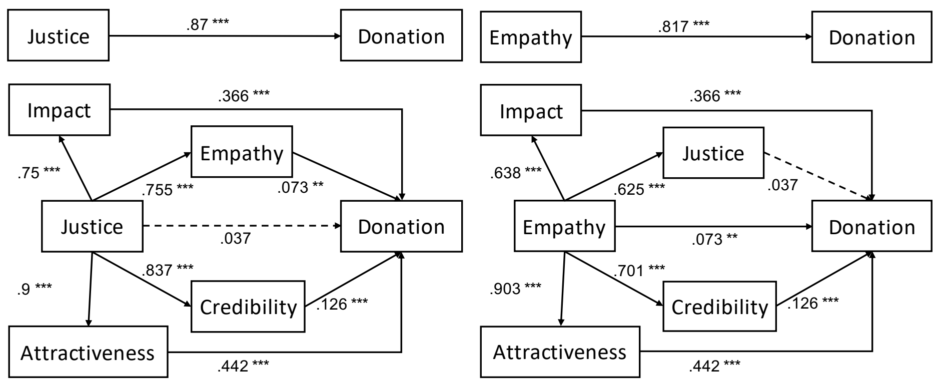
5.2 Regression Analysis for First Impressions and Donation Intention

To explore the relations between first impressions and the perceived donation intention (**RQ2**), we conduct regression analysis between first impressions (IVs) and the donation intention (DV). All IVs are first standardized by centering to zero mean and divided by two standard deviations following the pipeline in [41]. To ensure the feasibility of this method, we confirm that the pairwise Pearson correlation coefficients of predictor variables are all smaller than 0.8, which suggests the collinearity among first impressions is not severe [11]. We then calculate the variance inflation factor (VIF) among impression items; the resulting values are all smaller than 4, indicating that the multicollinearity issue does not exist [118].

We thus proceed to conduct regression analysis and list the results in Table 6. Model 1 to Model 5 analyzes the correlation between the individual dimensions of impression and the donation intention, respectively. These models suggest that all perceived impressions are positively correlated with viewers' inclinations to donate ($p < 0.001$), which is consistent with the role of these impressions manifested through multi-modality information in the donation crowdfunding area [131]. Model 6 combines all five impressions for regression analysis. We notice that the correlation coefficient (0.096) of perceived justice becomes insignificant in Model 6 (compared to 2.177 in Model 3), while the significant effect of the other impressions remains the same (Table 6). This may imply that perceived justice through an image is not a particularly powerful predictor [57] for the donation intention in the context of OMCCs. Model 7 removes the perceived justice and achieves similar performance to Model 6; all remaining impressions – perceived

Table 6: Regression models for impressions and donation intention. ***: p-value < 0.001.

Predictors	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Empathy	2.249 ***					0.215 ***	0.226 ***
Credibility		2.318 ***				0.352 ***	0.395 ***
Justice			2.177 ***			0.096	
Impact				2.423 ***		0.924 ***	0.946 ***
Attractiveness					2.562 ***	1.387 ***	1.398 ***
Intercept	4.538 ***	4.538 ***	4.538 ***	4.538 ***	4.538 ***	4.538 ***	4.538 ***
R^2	0.475	0.504	0.445	0.551	0.616	0.685	0.685



(a) Mediation analysis for perceived justice.

(b) Mediation analysis for perceived empathy.

Figure 4: Mediation analysis on how other first impressions affect the correlation between predictor and donation intention. This figure demonstrates the correlation of the selected predictor and donation intention before and after adding other first impressions. The solid line stands for $p < 0.05$, and the dash line stands for $p > 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

empathy, credibility, impact, and attractiveness – stay significantly correlated with potential donation decisions.

To gain a deeper understanding of the relation between different impressions, we further check the mediation effect among them via bootstrapped mediation analysis [49]. Figure 4(a) depicts the mediation effects of other first impressions for perceived justice toward the donation intention. It shows that the perceived justice is fully mediated by the other IVs ($p < 0.001$). This result is consistent with the regression analysis in Model 6 and indicates that the impression of justice based on a given image is not completely independent of other first impressions. Meanwhile, as illustrated in Figure 4(b), the perceived empathy is only partially mediated by the perceived credibility, impact, and attractiveness ($p < 0.001$), while the no mediation effect with significance is observed through the path of perceived justice. We obtain similar mediation analysis results for the perceived credibility, impact, and attractiveness and summarize them in Figure 10 in Section A.1.

In summary, we confirm the possible connection between first impressions created by OMCC images and viewers' tendency to donate. All dimensions of first impressions positively correlate with donation intention with significance. However, the perceived justice is not a particular powerful predictor for donation intention compared with empathy, credibility, impact, and attractiveness, as suggested by regression and mediation analysis. Since viewers can form a consistent perception of justice, we still analyze its correlation with visual features and whether it can be computationally modeled together with other impressions in the subsequent subsections. Meanwhile, we calculate the average rating across five annotators as the label of perceived first impressions for each image in the subsequent analysis and modeling.

5.3 Feature Analysis for Perceived First Impressions

We then seek answers for **RQ3**, what visual features of an OMCC image might be associated with the perceived empathy, credibility, justice, impact, and attractiveness. We compute the Spearman correlation between each feature and the five impression dimensions following [124] to understand their relations, if any. Note that for features related to human attributes such as age, gender, emotion, and face, we only include images with at least one recognized human face for the subsequent analysis.

Figure 5 illustrates the Spearman correlation between a specific feature (a column) and the five perceived first impressions each in a separate row. We only present the features that show significant correlation ($p < 0.05$) with at least one first impression. In the subsequent sections, we analyze and discuss the relationship between the most salient features shown in Figure 5. Table 7 lists the number and percentage of features in each set that significantly correlate with each type of first impression ($p < 0.05$). Overall, all four feature sets contain features that show a significant correlation with first impressions. Among the four feature sets, a majority (all $> 50\%$) of the content-based features correlate significantly with first impression ratings whereas the percentages in other sets are below 30%, indicating that the content of OMCC images closely relates to the crowd ratings on first impressions. This result is in

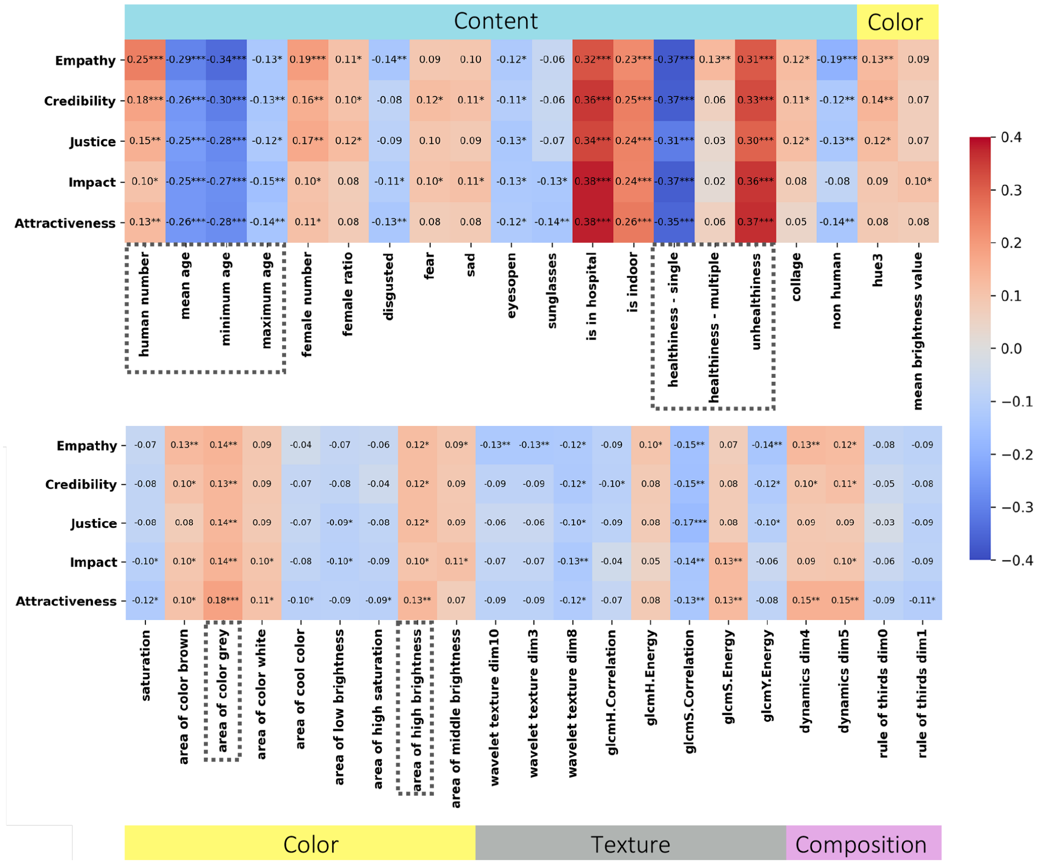


Figure 5: Correlation between image visual features and perceived first impressions. The value in each cell stands for the Spearman correlation coefficient, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. We elaborate on the possible connection between visual features in the dashed box and perceived first impressions in the subsequent sections.

Table 7: Number and ratios of significantly correlated features in each feature set ($p < 0.05$).

Feature set	Feature number	$p < 0.05$ number / percent				
		Empathy	Credibility	Justice	Impact	Attractiveness
Content-based	25	15 / 60%	15 / 60%	13 / 52%	14 / 56%	13 / 52%
Color-based	27	5 / 18.5%	4 / 14.8 %	4 / 14.8 %	8 / 29.6 %	7 / 25.9%
Texture-based	27	6 / 22.2%	4 / 14.8 %	3 / 11.1%	3 / 11.1%	3 / 11.1%
Composition-based	13	2 / 15.4%	2 / 15.4%	0 / 0	1/ 7.7%	3 / 23.1%
Full	92	28 / 30.4%	25 / 27.2%	20 / 21.7%	26 / 28.3%	26 / 28.3%

accordance with previous studies that viewers form the first impressions of OMCCs primarily based on the narrative elements [62], which emphasizes the necessity of presenting appropriate content in images of OMCCs. In particular, all age-related features show significant negative correlations with all first impressions. A previous study shows that viewers are more generous in donating to children compared with adults in the context of OMCCs [98], and our work further reveals the possible connection between age-related differences and the first impressions conveyed, implying that adults may need to exploit other elements in the image to boost viewers' first impressions.

The image theme feature provides a high-level description of an image, and the correlation between the theme and first impression ratings varies across image themes. Specifically, the unhealthiness narrative shows a strong positive correlation with all first impressions ($p < 0.001$). In contrast, healthiness narrative with a single person negatively correlates with all first impressions, although this is the most frequently appearing theme in our scraped OMCCs (45.2% of our scraped OMCCs). Figure 6(a) is an example of the image with the theme of healthiness narrative with a single person that receives low ratings on all dimensions of first impressions. Raters explain their low perceived first impressions given this image by *"the picture does not spark curiosity or sympathy"*, *"does not convey much context ... **unmotivated** to learn the situation"*. However, showing multiple people in healthiness narrative positively correlates with the perceived empathy ($p < 0.01$) and has no significant relations with the other impressions. Considering the strong positive connection between first impressions and

(a) An example of healthiness narrative with a single person receives low ratings on all dimensions of first impressions.

(b) An example of healthiness narrative with multiple people receives high ratings on all dimensions of first impressions.

(c) An example of healthiness narrative with a single person and bright background receives high ratings on the dimension of perceived credibility, justice, and attractiveness.

Figure 6: Samples of images rating. The value after E. (Empathy), C. (Credibility), J. (Justice), I. (Impact), A. (Attractiveness) stands for the received rating on that dimension, where the color purple represents "below average", and the color green represents "above average". Images © GoFundMe.

the donation intention, the different impacts on first impressions caused by image themes are in line with a previous finding that an unhealthiness narrative of the patient is more effective in attracting donation compared to a healthiness one [98]. Our result further suggests that including more people in the image could alleviate such a negative effect of the healthiness narrative to some extent. Figure 6(b) is an instance of an image with a healthiness narrative that receives high first impression ratings on all dimensions. One rater explains the rationale behind the scores, *"They appear to have a **strong male-to-male relationship**, perhaps father and son. Vulnerable men are underrated and more men should seek help."*

Consistent with the result from prior research that color can affect the donation intention [21], color-based features in our analysis also display a strong connection with ratings of first impressions. In particular, brightness positively correlates with perceived first impressions in all dimensions with significance. According to the theory of color psychology, higher brightness in the image could evoke lively emotions in humans [125] For example, Figure 6(c) is an example image of the healthiness narrative with a single person with a bright background; it receives high ratings on the dimension of perceived credibility, justice, and attractiveness. One rater explains the relatively high ratings for first impressions by "*I **enjoyed the bright pink shiny colors** and her smile*".

In summary, the correlation analysis shows that the OMCC image's theme has a strong association with the perceived first impressions, and an image of healthiness narrative with single people tends to get lower first impression ratings. However, multiple other visual features show a positive correlation with first impressions, such as displaying more people or having a large area of high brightness. However, the statistical correlation between first impressions and these visual features does not necessarily suggest causality. We denote it as one of the limitations in our work, and in the future, we will conduct causal experiments to ensure if causal effect actually exists. A possible setting is to conduct the in-situ photography processing on the original image, such as tuning the brightness and the color theme of the image, and compare whether the perceived first impressions would change compared with the original one through crowdsourcing.

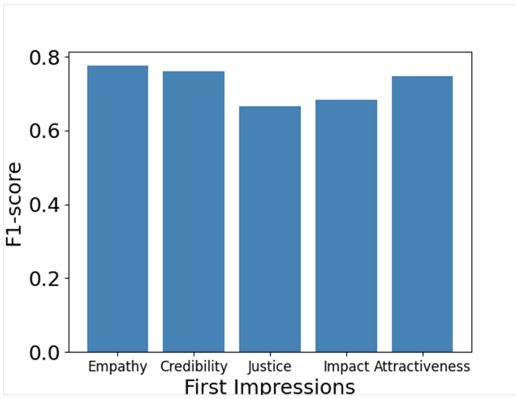
5.4 Supervised Prediction Performance

This section examines the possibility of predicting viewers' first impressions of OMCC images with computational models (**RQ4**). As demonstrated in Section 5.1, there exist controversial ratings for a few data. For example, one image receives ratings of $\{1, 2.67, 4, 7, 7\}$ (mean: 4.33, SD: 2.38) from five crowd workers on the dimension of attractiveness. This image will be classified as "low attractive" according to our criteria (i.e., the mean score is less than the mean of overall ratings on that dimension), which could be a controversial label since two out of five crowd workers rate 7 for it. Therefore, the average operation for conflicting ratings might introduce unreliable ground truth in the dataset, whereas the reliable ground truth is the prerequisite for training machine learning models [40]. To avoid machine learning models being affected by the controversial data, we filter out images with great inconsistent ratings (the standard deviation in any dimension is above the top 2.5%, as illustrated in Table 5) by following processing procedures in literature [39, 97, 127]. In total, 45 images are removed, most of which (53.3%) are in the theme of healthiness narrative with a single person. Since the number of initially sampled images under that theme is the largest, diverse images under that theme are kept for training classifiers after this step. Finally, 405 images are kept for computational modeling.

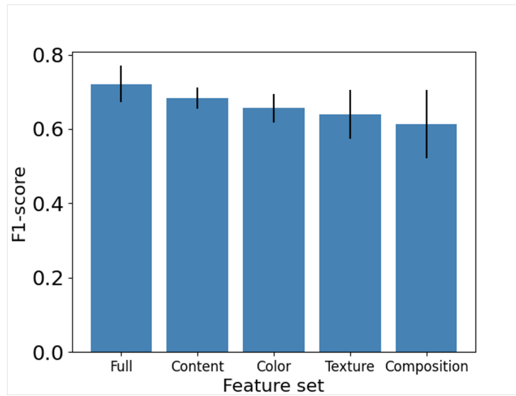
Table 8 summarizes the mean and standard deviation of precision, recall, and F1-score achieved by each classifier over five dimensions of first impressions on the test set. All our trained computational models outperform the random guess baseline by a large margin, demonstrating that viewers' first impressions of images in OMCCs are predictable. The distance-based model KNN achieves the lowest performance in terms of average recall (Mean: 0.632, SD: 0.06), which indicates that the KNN models fail to recognize certain images with relatively high perceived first impressions when trying to predict labels by matching features to those of the training samples. This result suggests that not all images that manifest great first impressions are alike, which aligns with previous work that potential donors have tolerance on images of OMCCs to some extent [62]. The non-linear kernel-based model SVC and MLP as well as the Decision Tree model improve the average recall and F1-score vastly, reflecting that these models have a higher ability to capture the

Table 8: The overall performance of classifiers. The "Mean" and "SD" represent the mean and standard deviation of the classifier performance across five dimensions of first impressions.

Model	Precision (Mean ± SD)	Recall (Mean ± SD)	F1 (Mean ± SD)
Random Guess	0.548 ± 0.044	0.497 ± 0.099	0.518 ± 0.065
KNN	0.669 ± 0.036	0.632 ± 0.06	0.648 ± 0.029
SVC	0.652 ± 0.102	0.673 ± 0.088	0.658 ± 0.076
MLP	0.676 ± 0.03	0.682 ± 0.029	0.679 ± 0.024
Decision Tree	0.652 ± 0.04	0.704 ± 0.098	0.673 ± 0.041
ResNet-34	0.677 ± 0.021	**0.763 ± 0.07**	0.711 ± 0.037
Random Forest	**0.733 ± 0.024**	0.723 ± 0.081	**0.727 ± 0.049**



(a) Evaluation on each impression.



(b) Evaluation on each feature set.

Figure 7: Evaluation of the Random Forest Model. Figure 7(a) depicts the F1-score on each dimension of first impressions, and Figure 7(b) illustrates the mean and standard deviation of F1-score across five dimensions of first impressions.

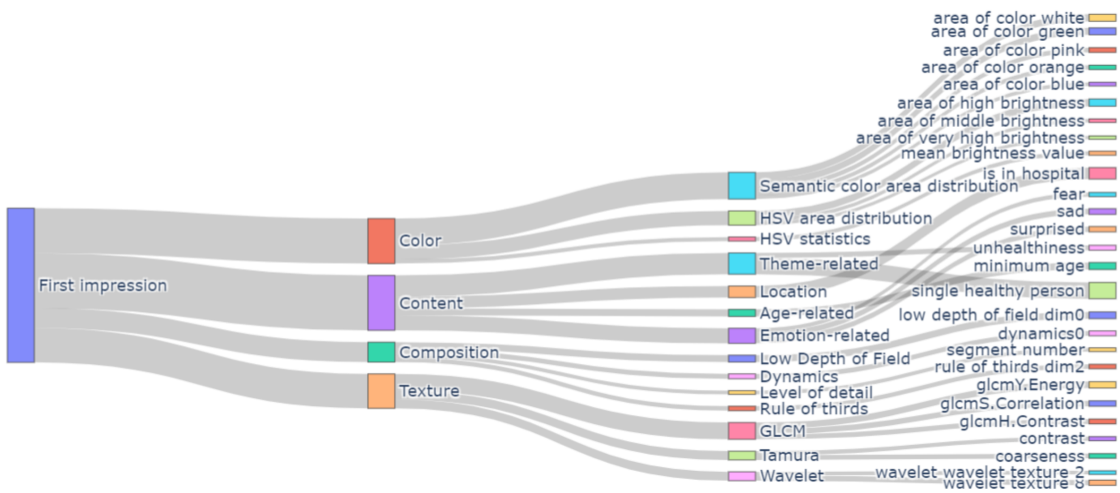
relation between image features and perceived impressions. The best machine learning model, Random Forest (RF), further boosts the classification performance on all three measures. A possible reason is that the RF model ensembles multiple decision trees , and is less sensitive to the hyper-parameters when applied in a variety of tasks [2]. The result is similar to a previous study on modeling the perceived impression of mobile app screenshots [127], where the RF model performs the best, showing the generalizability of ensemble trees. In addition, RF outperforms the deep-learning-based model ResNet-34, on the overall precision and F1-score, which shows the effectiveness of our extracted and manually annotated image features. Although the ResNet-based model has been successfully applied to multiple scenarios via transfer learning, such as scene classification [75] and cell image classification [96], these models generally require enumerating training samples for decent performance (typically over 10 thousand images), which is much larger than our dataset. In summary, the perceived impressions can be computationally modeled with a machine learning model based on our extracted and manually annotated visual features.

Then we employ and re-train the Random Forest model to predict individual impressions. Figure 7(a) presents the F1-score of the RF model achieved on each dimension of first impressions. The model achieves better performance on the dimension of empathy (F1-score: 0.776), credibility (F1-score: 0.761), and attractiveness (F1-score: 0.747), while it has a slightly lower F1-score on predicting perceived impact (F1-score: 0.682) and justice (F1-score: 0.667). The result implies that the proposed features have relatively lower prediction power on the perceived impact and justice. A possible reason is that the perceived impact and justice could be affected by the dissimilar socio-economic backgrounds of viewers [70, 85]. Thus our extracted and manually annotated features may not be suitable for capturing the first impressions on those two dimensions. In the future, we could combine more information for better classification performance.

In summary, we demonstrate that the first impressions of images in OMCCs can be computationally modeled, and the best model, Random Forest, achieves an average F1-score of 0.727. In the next section, we address **RQ5** by analyzing the contributing factors for prediction.

5.5 Analysis for Feature Importance in Prediction Model

To understand the contribution of different categories of features for predicting each type of impression (**RQ5**), we train RF models based on separate feature sets. Figure 7(b) compares the F1-scores of Random Forest models trained with different feature sets, and that trained with full features. By comparing models trained with a single feature set, the one using the content-based features achieves the highest performance (mean: 0.693, SD: 0.028), followed by those employing color-based features (mean: 0.663, SD: 0.038) and texture-based features (mean: 0.657, SD: 0.065); the one applying composition-based features (mean: 0.641, SD: 0.092) comes last. The relative predictive power of feature sets is consistent with the



(a) The relative importance of the main contributing factors to the final model in predicting impressions.

	Emapthy		Credibility		Justice		Impact		Attractiveness	
Rank	Feature name	Importance	Feature name	Importance	Feature name	Importance	Feature name	Importance	Feature name	Importance
1	single healthy person	0.044	single healthy person	0.051	single healthy person	0.051	single healthy person	0.059	single healthy person	0.044
2	is in hospital	0.033	is in hospital	0.030	is in hospital	0.039	is in hospital	0.044	area of color white	0.036
3	area of color green	0.028	minimum age	0.029	area of high brightness	0.030	low depth of field0	0.033	is in hospital	0.031
4	glcmH.Contrast	0.027	glcmS.Correlation	0.027	rule of thirds2	0.026	area of color white	0.032	glcmY.Energy	0.031
5	minimum age	0.025	surprised	0.024	minimum age	0.026	area of color pink	0.028	unhealthiness	0.029
6	low depth of field0	0.024	area of high brightness	0.023	low depth of field0	0.022	dynamics0	0.026	area of high brightness	0.028
7	glcmY.Energy	0.023	fear	0.023	contrast	0.020	mean brightness value	0.023	area of color green	0.027
8	area of middle brightness	0.022	area of color blue	0.021	coarseness	0.020	coarseness	0.019	sad	0.024
9	surprised	0.020	area of color orange	0.019	wavelet texture 2	0.019	sad	0.018	segment number	0.021
10	area of very high brightness	0.019	area of color green	0.019	wavelet texture 8	0.019	glcmY.Energy	0.017	wavelet texture 8	0.021

(b) The top 10 important features for predicting each dimension of first impressions.

Figure 8: The relative importance of the main contributing factors to the final model in predicting impressions

total number of features showing a significant correlation with first impressions in each set. This result highlights the importance of content-based features in establishing proper first impressions in OMCCs [62]. Nevertheless, the best performance is achieved by incorporating these different feature sets into the model, which shows the necessity of considering image features from multiple aspects for computationally predicting first impressions.

We further conduct Gini importance analysis [14] for visual features in the RF model to understand the relative contribution of individual features for classification. Figure 8(a) depicts the relative importance of the top 27 features in the RF model, with which the model obtains similar performance (an average of 0.711 precision, 0.689 recall, and 0.698 F1-score) to that of the full model. The top 10 contributing factors for modeling each type of the first impression are listed in Figure 8(b). In general, the content-based features have the highest total importance score across models for predicting five first impressions. Interestingly, the importance of the healthiness narrative with a single person ranks the highest among all prediction tasks, indicating that it is essential to select a proper theme for the OMCC image. The hospital scenery, photographee's minimum age, photographee's emotions, including being sad, fearful, and surprised, also seem to play a key role in the predictions. This is in line with findings in previous works that showing hospital scenery [123], photographee's age [82] and emotional delivery [99] could affect the performance of OMCCs. Some low-level visual features also have a considerable impact on the classifier's performance. For example, the area of the color green is among the top 10 important features for predicting perceived empathy, credibility, and attractiveness. Previous literature suggests that the color green closely connects with the perceived credibility of the website [104]. Meanwhile, the color green could invoke the viewer's emotion of hope and excitement, which associates with empathy [65] and attractiveness [117], respectively. Note that among these 27 crucial features for predicting first impressions, 14 of them (including the area of the color green) do not have a significant Spearman correlation with first impressions as shown in Figure 5. This may suggest the relationship between some image visual features and first impressions may not be monotonic in OMCC images [12].

In short, the content-based features show superior prediction power for first impressions modeling, while incorporating other feature sets further enhances the model performance. In particular, we highlight the importance of selecting a proper theme for the OMCC image. Moreover, we suggest there may exist a non-monotonic relationship between image visual features and first impressions in the context of OMCC.

6 DISCUSSION

6.1 Reflection on computational modeling for perceived first impressions

Figure 9: The sample (a) and (d) receive relatively high ratings on all dimensions of first impressions, whereas the sample (b) and (c) receives relative low ratings for all perceived first impressions. The prediction model successfully predicts the first impressions for the sample (a) and (b), however, fails for the sample (c) and (d). The ↑ / ↓ represent the classification result by the model ("above average" / "below average"), and the red arrow denotes error in prediction. Images © GoFundMe.

Figure 9 presents four examples with the ratings of crowd workers and the classification result by our RF models. Although our RF model achieves reasonable classification performance, it fails in some instances. Based on these error cases, we summarize two flaws in the model. First, our model may not fully understand the combinatorial relationship between features. For example, Figure 9(c) contains a single healthy adult on a vehicle in the outdoor scene. The model mistakenly assumes that the image delivers proper impressions on the dimension of perceived empathy, justice, and attractiveness. A possible reason could be that the grey area dominates the color of that image (the vehicle and ground part). Our correlation analysis in Section 5.3 shows that the area of the color grey positively correlates with all impressions ($p < 0.01$). However, in our annotated dataset, the large area of color grey often appears in black and white images (e.g., Figure 9(a)), while the black and white narrative is suggested to affect human emotions [36]. The failure prediction in Figure 9(c) implies that the model fails to understand the contribution of the area of color grey more due to the color theme rather than specific elements in the image.

Another bottleneck is some visual features are not included in our feature sets. For instance, Figure 9(d) is an underrepresented example that utilizes the text and symbol to demonstrate the unhealthiness of the man. Proper first impressions on all dimensions except for justice are delivered through this image. For example, one rater mentions *"Cancer is such a devastating illness, the grandfather "Pop" made me interested in more about him and his illness."*. However, the text and symbol-related features are not included in our feature set, and thus the model can not understand the semantic meaning of presenting this information.

Despite the possible failure in predicting first impressions from images, we do not claim that viewers' first impressions given the image are the same as those toward the campaign, as content could affect viewers' attitudes as well [68]. Nor do we assert that viewers' high perceived first impressions must be associated with high donation intention since the donation intention could be influenced by other factors like the donation progress of an OMCC [74]. And finally, a viewer's high donation intention does not necessarily imply a high donation amount, because the actual donation of a viewer could be affected by other issues, e.g., the lack of money [8]. However, our study demonstrates some elementary understanding of the connection between first impressions and donation intention of OMCC and the possible visual factors that may affect viewer's first impressions, and sheds light on modeling viewer's first impressions with visual features via computational approaches. We discuss how to apply the initial insights of this work to a feedback system for supporting fundraisers of OMCCs in the subsequent section.

6.2 Potential applications

6.2.1 Application for OMCCs. A potential application of our work is a feedback system integrated with OMCPs, which can provide an instant, automatic assessment of whether an uploaded image could deliver appropriate first impressions. As reported in [61], some fundraisers would worry about whether viewers could understand their financial needs based on their selected images. Our system could benefit these fundraisers by providing a mirror through which they can get to know viewers' judgment beforehand. The fundraiser can first prepare multiple candidate images for initiating an OMCC and then know how well these images convey the first impression of each dimension with the help of the system. Eventually, they pick up images that match the first impressions they expect to convey for OMCCs.

Meanwhile, the platforms can recognize important features in the system for predicting first impressions, especially those negatively correlated with first impressions. If the system identifies potential problems in the image provided by the fundraisers, the platform could remind fundraisers of the potential negative impact on viewers' first impressions. For example, images of healthiness narrative with a single person are the most frequently used image theme in our collected images. Fundraisers could be reminded of the negative effect of that image theme in delivering first impressions. In addition, the system can summarize features that have the potential to increase certain first impressions and present them to fundraisers as recommendations, which can function as modification direction and facilitate fundraisers finding proper images.

It is worth mentioning that the manually annotated image location-related features are strongly correlated with the viewers' first impressions with OMCC images, and play critical roles in predicting these impressions. In particular, the location feature of "is in hospital" is among the top-3 important features for predicting all five dimensions of first impressions. To make it practical to deploy image assessment in OMCPs, the timely identification of the location features needs to be implemented. One possible solution is to invite fundraisers to add location tags to the selected images. Another promising approach is automatically recognizing location features with the deep learning technique. For instance, OMCPs can fine-tune the pre-trained CNN model on the large-scale annotated location dataset, which is similar to our demonstrated pipeline of recognizing OMCC image theme with the ResNet-34 based classifier in Section 3.2. The location annotation can be obtained by engaging viewers to provide the location tags for the viewed OMCC images. Meanwhile, to guarantee the human tagging quality, OMCPs can refer to the protocol proposed by reCAPTCHA [121], which achieves a character annotation accuracy of over 99% by massive internet users. Eventually, manually annotating location features for OMCC images could be reliably automated.

6.2.2 Generalizability of Our Work. Although this work focuses on the visual information of OMCCs, content information also influences the first impressions of viewers a lot [68]. The content features of OMCCs can easily be incorporated into our model to consider the multi-modality effect simultaneously. Possible content-related features could include content statistic features (e.g., content length) and LIWC features, which are commonly employed by works on content analysis of online crowdfunding campaigns [92, 138]. The multi-modality effect can be captured by feeding both visual features and content features into the RF model.

Meanwhile, although the platform of GoFundMe typically contains only one cover image, other platforms like Qingsongchou, generally contain multiple cover images that could play complementary roles in manifesting first impressions [123]. Our approach can be generalized into these platforms with a cascaded architecture that first predicts the first impressions conveyed from every image using the RF model in this work, then combine another subsequent model to capture the mutual influence of first impression manifestation through multiple images.

Moreover, this work could be applied in other scenarios where analysis and prediction of visual impressions could help stakeholders or system users, such as visualization design [48] and video selection for commercial crowdfunding [30]. However, researchers should tailor the essential impressions, major media type, and interpretable features accordingly.

6.3 Ethical concerns

6.3.1 Potential misuse. Fraud exists in OMCCs through exaggerating, or even faking one's financial need [110]. It is possible that the feedback system may facilitate fraud through OMCCs, since people could intentionally choose images that can adequately manifest first impressions. To eliminate the deception in OMCCs for better allocation of social resources, we should pay more attention to recognizing fraud campaigns. First, OMCPs can detect the misuse patterns of fundraisers in selecting the OMCC images with the feedback system. OMCPs can first collect user behavior patterns of using the system, and the label of whether an OMCC is a fraud or not, then automatically identify malicious users and associated behaviors via data-driven approaches [109, 129]. Once the system recognizes potential misuse of the feedback system, the platform should further examine these OMCCs before they get launched.

Second, since fraud OMCCs exist even if without the feedback system, the platforms should track and carefully examine the validity of campaigns after they are launched. A possible solution is to develop a fraud detection algorithm that can utilize multi-modality information, such as images and texts, fundraiser's profile information, and the sharing related features. Another direction could be interface design. As suggested by [103], a novel interface that shows the reputation data of sellers can support customers recognizing fraud in e-commerce. Therefore, platforms can design the interface that presents credibility-related information of the campaign and the fundraiser, which may facilitate viewers recognizing fraud OMCCs to prevent them from being cheated. In addition, as recommended by Zenone et al. [133], the platform could develop protocols to verify the fundraiser's identity, and make it easier for viewers to report fraud.

6.3.2 Ethical implications. To deploy the first impression feedback model into a real application, several ethical issues should be considered. First, the system developer should consider the potential bias issue in the model. On the one hand, the training data for the model should cover diverse images. Otherwise, the model may not provide accurate predictions for less common cases. It would be essential for underrepresented people, for whom current algorithms rarely obtain decent performance in recognizing their images [10]. On the other hand, certain visual factors, e.g., gender-related features, are not suggested for explaining the predicted first impressions to fundraisers. It is possible to frustrate fundraisers as such explanations could leave them a sense of discrimination [46].

Second, fundraisers' privacy should be protected. Fundraisers often feel guilty in seeking financial support, and do not hope to become the focus over the internet [90]. Therefore, the feedback system should not recommend images in other OMCCs that convey strong first impressions to another fundraiser. Such an interaction would both offense others' privacy and possibly make the fundraiser raise privacy concerns in using the system.

Third, the evaluation feedback should be diverse. Similar to the phenomenon that websites are becoming similar caused by color scheme standardization and overlap in source codes [45], images for OMCCs are probably to become alike if similar suggestions on first impression management are provided to multiple users continuously. The homogenization of OMCC images could decrease the perceived attractiveness of these images as viewers are more engaged to unconventional things [86]. Moreover, repeatedly appearing similar images could cause viewers to doubt the sincerity of OMCCs [64]. However, sincerity is crucial in the context of OMCCs, as it closely connects with the perceived empathy [15] and credibility [54]. Eventually, the system may lose the efficacy of impression management caused by the homogenization of images. Meanwhile, the system developer should collect feedback of first impressions from viewers regularly, and adopt the online learning mechanism [4] to update the model and recommendations accordingly, to ensure the feedback remains effective.

6.4 Limitations and Future Works

This work contains several limitations. First, we restrain all the annotated OMCCs and viewers in the US. As the culture could influence the first impressions greatly [71], the same conclusion may not apply to OMCCs in other countries. Future work could explore the difference in perceived first impressions across countries. Second, our work does not distinguish the difference between viewers and provides a general viewers prediction result. However, the demographic information (e.g., age, gender) and personality of a viewer (e.g., agreeableness) have the potential to influence the perceived first impressions [81, 83]. In the future, we will incorporate the user demographic information and user profile to boost the classifier's performance further. Third, we notice that controversial ratings exist for certain images in Section 5.4. Follow-up work could be investigating why viewers have controversial perceptions and what image elements might cause the difference. Lastly, as mentioned in Section 5.3, we provide a correlational analysis between first impressions and interpretable visual features, instead of conducting causal experiments. In the future, we would first explore the casual connection between visual features and perceived first impressions. Based on the identified causal relations, it could be promising to design photography processing tools for OMCCs, e.g., color filtering [128], which provides an in-situ modification of the image to facilitate the manifestation of first impressions.

7 CONCLUSION

In this paper, we present our understanding of viewers' first impressions (i.e., perceived empathy, credibility, justice, impact, and attractiveness) with images in online medical crowdfunding campaigns (OMCCs), and explore the modeling of these first impressions via data-driven methods. Based on our crowdsourced data, we confirm that viewers can establish substantial agreement on all five dimensions of first impressions and donation intention with OMCC images. Moreover, viewers' first impressions are positively correlated with the projected donation intentions. We further predict whether an OMCC image could deliver proper first impressions by feeding extracted and manually annotated visual features (i.e., image content features, color features, texture features, and composition features) into machine learning models. The results suggest that our classifiers can achieve reasonable high performance for first impressions modeling. Combined with the correlation analysis between visual features and perceived first impressions, crowd workers' rationale for ratings, and contributing factors for predicting first impressions, we identify essential factors for manifesting first impressions, such as the image's theme and brightness. Our work could benefit fundraisers preparing OMCCs by automatically assessing whether the selected images could deliver appropriate first impressions to viewers.

ACKNOWLEDGMENTS

Many thanks to the anonymous reviewers for their insightful suggestions. We thank Juanru Fang, Shixu Zhou, and Jindong Han for their valuable inputs. This research was supported in part by HKUST-WeBank Joint Laboratory Project Grant No.: WEB19EG01-d.

REFERENCES

[1] Farida B Ahmad and Robert N Anderson. 2021. The leading causes of death in the US for 2020. *JAMA* 325, 18 (2021), 1829–1830.

[2] Jehad Ali, Rehanullah Khan, Nasir Ahmad, and Imran Maqsood. 2012. Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)* 9, 5 (2012), 272.

[3] Manuel Alonso Dos Santos, Camila Lobos, Nathalie Muñoz, Dámaris Romero, and Ricardo Sanhueza. 2017. The influence of image valence on the attention paid to charity advertising. *Journal of Nonprofit & Public Sector Marketing* 29, 3 (2017), 346–363.

[4] Terry Anderson. 2008. *The theory and practice of online learning*. Athabasca University Press.

[5] Howard Barich and Philip Kotler. 1991. A framework for marketing image management. *MIT Sloan Management Review* 32, 2 (1991), 94.

[6] C Daniel Batson. 1987. Prosocial motivation: Is it ever truly altruistic? *Advances in experimental social psychology* 20 (1987), 65–122.

[7] C Daniel Batson, Tricia R Klein, Lori Highberger, and Laura L Shaw. 1995. Immorality from empathy-induced altruism: When compassion and justice conflict. *Journal of personality and social psychology* 68, 6 (1995), 1042.

[8] René Bekkers and Pamala Wiepking. 2011. Who gives? A literature review of predictors of charitable giving part one: Religion, education, age and socialisation. *Voluntary Sector Review* 2, 3 (2011), 337–365.

[9] Paul Belleflamme, Thomas Lambert, and Armin Schwienbacher. 2014. Crowdfunding: Tapping the right crowd. *Journal of business venturing* 29, 5 (2014), 585–609.

[10] Cynthia L Bennett, Cole Gleason, Morgan Klaus Scheuerman, Jeffrey P Bigham, Anhong Guo, and Alexandra To. 2021. "It's Complicated": Negotiating Accessibility and (Mis) Representation in Image Descriptions of Race, Gender, and Disability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19.

[11] William D Berry, Stanley Feldman, and Dr Stanley Feldman. 1985. *Multiple regression in practice*. Number 50. Sage.

[12] Douglas G Bonett and Thomas A Wright. 2000. Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika* 65, 1 (2000), 23–28.

[13] Douglas G Bonett and Thomas A Wright. 2015. Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of organizational behavior* 36, 1 (2015), 3–15.

[14] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. 2017. *Classification and regression trees*. Routledge.

[15] Sabrina Brugel, Marie Postma-Nilsenová, and Kiek Tates. 2015. The link between perception of clinical empathy and nonverbal behavior: The effect of a doctor's gaze and body orientation. *Patient education and counseling* 98, 10 (2015), 1260–1265.

[16] Gordon Burtch and Jason Chan. 2018. Investigating the relationship between medical crowdfunding and personal bankruptcy in the United States: Evidence of a digital divide. *MIS Quarterly (Forthcoming)* (2018).

[17] C Daryl Cameron and B Keith Payne. 2011. Escaping affect: how motivated emotion regulation creates insensitivity to mass suffering. *Journal of personality and social psychology* 100, 1 (2011), 1.

[18] Rich Caruana, Steve Lawrence, and Lee Giles. 2001. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Advances in neural information processing systems* (2001), 402–408.

[19] CDC. 2021. Main Types of Stroke. https://www.cdc.gov/stroke/types_of_stroke.htm.

[20] En-Chi Chang and Ya-Fen Tseng. 2013. Research note: E-store image, perceived value and perceived risk. *Journal of business research* 66, 7 (2013), 864–870.

[21] Jungsil Choi, Yexin Jessica Li, Priyamvadha Rangan, Bingqing Yin, and Surendra N Singh. 2020. Opposites attract: Impact of background color on effectiveness of emotional charity appeals. *International Journal of Research in Marketing* 37, 3 (2020), 644–660.

[22] Domenic V Cicchetti. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment* 6, 4 (1994), 284.

[23] Jay S Coke, C Daniel Batson, and Katherine McDavis. 1978. Empathic mediation of helping: a two-stage model. *Journal of personality and social psychology* 36, 7 (1978), 752.

[24] Joe Cox, Thang Nguyen, Andy Thorpe, Alessio Ishizaka, Salem Chakhar, and Liz Meech. 2018. Being seen to care: The relationship between self-presentation and contributions to online pro-social crowdfunding campaigns. *Computers in Human Behavior* 83 (2018), 45–55.

[25] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. 2006. Studying aesthetics in photographic images using a computational approach. In *European conference*

on computer vision. Springer, 288–301.
[26] Frans BM De Waal. 2008. Putting the altruism back into altruism: the evolution of empathy. *Annu. Rev. Psychol.* 59 (2008), 279–300.
[27] Jean Decety and Philip L Jackson. 2004. The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews* 3, 2 (2004), 71–100.
[28] Kim Demandt. 2019. *Why do people donate to online health crowdfunding?* B.S. thesis. University of Twente.
[29] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.
[30] Sanorita Dey, Brittany Duff, Karrie Karahalios, and Wai-Tat Fu. 2017. The art and science of persuasion: not all crowdfunding campaign videos are the same. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 755–769.
[31] Stephan Dickert, Janet Kleber, Daniel Västfjäll, and Paul Slovic. 2016. Mental imagery, impact, and affect: A mediation model for charitable giving. *PloS one* 11, 2 (2016), e0148274.
[32] Stephan Dickert, Namika Sagara, and Paul Slovic. 2011. Affective motivations to help others: A two-stage model of donation decisions. *Journal of Behavioral Decision Making* 24, 4 (2011), 361–376.
[33] Paul Doncaster. 2014. *The UX Five-Second Rules: Guidelines for User Experience Design's Simplest Testing Technique*. Morgan Kaufmann.
[34] Richard Wayne Dosselmann and Xue Dong Yang. 2011. *No-reference image quality assessment using level-of-detail*. Citeseer.
[35] Nancy Eisenberg and Paul A Miller. 1987. The relation of empathy to prosocial and related behaviors. *Psychological bulletin* 101, 1 (1987), 91.
[36] Andrew J Elliot and Markus A Maier. 2014. Color psychology: Effects of perceiving color on psychological functioning in humans. *Annual review of psychology* 65 (2014), 95–120.
[37] Arvid Erlandsson, Fredrik Björklund, and Martin Bäckström. 2015. Emotional reactions, perceived impact and perceived responsibility mediate the identifiable victim effect, proportion dominance effect and in-group effect respectively. *Organizational Behavior and Human Decision Processes* 127 (2015), 1–14.
[38] Chris D Frith and Uta Frith. 2006. The neural basis of mentalizing. *Neuron* 50, 4 (2006), 531–534.
[39] Yujia Gao, Jinu Jang, and Diyi Yang. 2021. Understanding the Usage of Online Media for Parenting from Infancy to Preschool At Scale. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.
[40] Enrique Garcia-Ceja, Michael Riegler, Tine Nordgreen, Petter Jakobsen, Ketil J Oedegaard, and Jim Tørresen. 2018. Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing* 51 (2018), 1–26.
[41] Andrew Gelman. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in medicine* 27, 15 (2008), 2865–2873.
[42] GoFundMe. 2010. GoFundMe: No.1 Fundraising Platform for Crowdfunding. https://www.gofundme.com/.
[43] GoFundMe. 2021. These Valuable Facebook Fundraising Tips Can Boost Donations. https://www.gofundme.com/c/blog/facebook-fundraising.
[44] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT press.
[45] Samuel Goree, Bardia Doosti, David Crandall, and Norman Makoto Su. 2021. Investigating the Homogenization of Web Design: A Mixed-Methods Approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
[46] David Hankerson, Andrea R Marshall, Jennifer Booker, Houda El Mimouni, Imani Walker, and Jennifer A Rode. 2016. Does technology have race?. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 473–486.
[47] Robert M Haralock and Linda G Shapiro. 1991. *Computer and robot vision*. Addison-Wesley Longman Publishing Co., Inc.
[48] Lane Harrison, Katharina Reinecke, and Remco Chang. 2015. Infographic aesthetics: Designing for the first impression. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1187–1190.
[49] Andrew F Hayes. 2009. Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication monographs* 76, 4 (2009), 408–420.
[50] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
[51] United Healthcare. 2021. Critical illness insurance. https://www.uhc.com/dental-vision-supplemental-plans/critical-illness-insurance.
[52] Martin L Hoffman. 1990. Empathy and justice motivation. *Motivation and emotion* 14, 2 (1990), 151–172.
[53] Martin L Hoffman. 2001. *Empathy and moral development: Implications for caring and justice*. Cambridge University Press.
[54] Niek Hoogervorst, Judith Metz, Lonneke Roza, and Eva van Baren. 2016. How perceptions of altruism and sincerity affect client trust in volunteers versus paid workers. *Nonprofit and Voluntary Sector Quarterly* 45, 3 (2016), 593–611.
[55] Jian-Ren Hou, J. Zhang, and Kunpeng Zhang. 2019. Can Title Images Predict the Emotions and the Performance of Crowdfunding Projects?. In *HICSS*.
[56] Jian-Ren Hou, Jie Jennifer Zhang, and Kunpeng Zhang. 2020. Pictures that are Worth a Thousand Donations: How Emotions in Project Images Drive the Success of Crowdfunding Campaigns? An Image Design Perspective. *An Image Design Perspective (October 20, 2020)* (2020).
[57] Yiqing Hua, Mor Naaman, and Thomas Ristenpart. 2020. Characterizing twitter users who engage in adversarial interactions against political candidates. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
[58] Nurulhuda Ibrahim, Kok Wai Wong, and Mohd Fairuz Shiratuddin. 2015. Persuasive impact of online media: investigating the influence of visual persuasion. In *2015 Asia Pacific Conference on Multimedia and Broadcasting*. IEEE, 1–7.
[59] National Cancer Institute. 2020. Cancer Statistics for 2020. https://www.cancer.gov/about-cancer/understanding/statistics.
[60] Jungseock Joo, Weixin Li, Francis F Steen, and Song-Chun Zhu. 2014. Visual persuasion: Inferring communicative intents of images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 216–223.
[61] Jennifer G Kim, Hwajung Hong, and Karrie Karahalios. 2018. Understanding identity presentation in medical crowdfunding. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.
[62] Jennifer G Kim, Ha Kyung Kong, Karrie Karahalios, Wai-Tat Fu, and Hwajung Hong. 2016. The power of collective endorsements: credibility factors in medical crowdfunding campaigns. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 4538–4549.
[63] Jennifer G Kim, Kristen Vaccaro, Karrie Karahalios, and Hwajung Hong. 2017. " Not by Money Alone" Social Support Opportunities in Medical Crowdfunding Campaigns. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1997–2009.
[64] Taewook Kim, Jung Soo Lee, Zhenhui Peng, and Xiaojuan Ma. 2019. Love in Lyrics: An Exploration of Supporting Textual Manifestation of Affection in Social Messaging. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–27.
[65] Young-Hee Kim and Tae-Soon Koh. 2014. Relationships among Hope, Empathy and Happiness of Middle School Student. *The Journal of the Korea Contents Association* 14, 10 (2014), 330–338.
[66] Nicholas J Klein, Minh Tran, Sarah Riley, et al. 2020. Desperately need a car": Analyzing crowdfunding campaigns for car purchases and repairs on Gofundme.com. *Travel Behaviour and Society* 21 (2020), 247–256.
[67] Terry K Koo and Mae Y Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine* 15, 2 (2016), 155–163.
[68] DongIl Lee and JaeHong Park. 2018. How Content Features of Charity Crowdfunding Projects Attract Potential Donors?. In *Workshop on E-Business*. Springer, 124–131.
[69] Saerom Lee, Karen Page Winterich, and William T Ross Jr. 2014. I'm moral, but I won't help you: The distinct roles of empathy and justice in donations. *Journal of Consumer Research* 41, 3 (2014), 678–696.
[70] Stefan Liebig, Peter Valet, and Jürgen Schupp. 2010. Perceived income justice depends on the economy. *Weekly Report* 6, 24 (2010), 187–192.
[71] Gitte Lindgaard, Justyna Litwinska, and Cathy Dudek. 2008. Judging web page visual appeal: Do east and west really differ. *Proc. IADIS* 8 (2008), 157–164.
[72] Jianli Liu, Edwin Lughofer, and Xianyi Zeng. 2015. Aesthetic perception of visual textures: a holistic exploration using texture analysis, psychological experiment, and perception modeling. *Frontiers in computational neuroscience* 9 (2015), 134.
[73] Lili Liu, Ayoung Suh, and Christian Wagner. 2017. Donation behavior in online micro charities: An investigation of charitable crowdfunding projects. In *Proceedings of the 50th Hawaii International Conference on system sciences*.
[74] Lili Liu, Ayoung Suh, and Christian Wagner. 2018. Empathy or perceived credibility? An empirical study on individual donation behavior in charitable crowdfunding. *Internet Research* (2018).
[75] Shaopeng Liu, Guohui Tian, and Yuan Xu. 2019. A novel scene classification model combining ResNet based transfer learning and data augmentation with a filter. *Neurocomputing* 338 (2019), 191–206.
[76] Marsha D Loda, Karin Teichmann, and Andreas H Zins. 2009. Destination websites' persuasiveness. *International journal of culture, tourism and hospitality research* (2009).
[77] Jana Machajdik and Allan Hanbury. 2010. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the 18th ACM international conference on Multimedia*. 83–92.
[78] Rojas Q Mario, David Masip, and Jordi Vitrià. 2011. Predicting dominance judgements automatically: a machine learning approach. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 939–944.
[79] Carolyn McClanahan. 2018. People Are Raising 650 Million On GoFundMe Each Year To Attack Rising Healthcare Costs. https://www.forbes.com/sites/carolynmcclanahan/2018/08/13/using-gofundme-to-attack-health-care-costs/
[80] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282.

[81] Barbara McLellan and Stuart J McKelvie. 1993. Effects of age and gender on perceived facial attractiveness. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 25, 1 (1993), 135.

[82] Paul Messaris. 1997. *Visual persuasion: The role of images in advertising*. Sage.

[83] Todd A Mooradian, Mark Davis, and Kurt Matzler. 2011. Dispositional empathy and the hierarchical structure of personality. *The American journal of psychology* 124, 1 (2011), 99–109.

[84] Pere Obrador, Ludwig Schmidt-Hackenberg, and Nuria Oliver. 2010. The role of image composition in image aesthetics. In *2010 IEEE International Conference on Image Processing*. IEEE, 3185–3188.

[85] Garrett J O'Keefe, Kathaleen Nash, and Jenny Liu. 1981. The perceived utility of advertising. *Journalism Quarterly* 58, 4 (1981), 535–542.

[86] Andrew Ortony, Gerald L Clore, and Allan Collins. 1990. *The cognitive structure of emotions*. Cambridge university press.

[87] Charles Egerton Osgood, George J Suci, and Percy H Tannenbaum. 1957. *The measurement of meaning*. Number 47. University of Illinois press.

[88] Paula Otero. 2015. Crowdfunding. A new option for funding health projects. *Arch Argent Pediatr* 113, 2 (2015), 154–157.

[89] Antti Oulasvirta, Samuli De Pascale, Janin Koch, Thomas Langerak, Jussi Jokinen, Kashyap Todi, Markku Laine, Manoj Kristhombuge, Yuxi Zhu, Aliaksei Miniukovich, et al. 2018. Aalto Interface Metrics (AIM) A Service and Codebase for Computational GUI Evaluation. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*. 16–19.

[90] Varsha Palad and Jeremy Snyder. 2019. "We don't want him worrying about how he will pay to save his life": Using medical crowdfunding to explore lived experiences with addiction services in Canada. *International Journal of Drug Policy* 65 (2019), 73–77.

[91] Lee-Yun Pan and Jyh-Shen Chiou. 2011. How much can you trust online information? Cues for perceived trustworthiness of consumer-generated online information. *Journal of interactive marketing* 25, 2 (2011), 67–74.

[92] Annaleena Parhankangas and Maija Renko. 2017. Linguistic style and crowdfunding success among social and commercial entrepreneurs. *Journal of business venturing* 32, 2 (2017), 215–236.

[93] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019), 8026–8037.

[94] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830.

[95] QingSongChou. 2014. qschou. https://qschou.com/.

[96] A Sai Bharadwaj Reddy and D Sujitha Juliet. 2019. Transfer learning with ResNet-50 for malaria cell-image classification. In *2019 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, 0945–0949.

[97] Katharina Reinecke, Tom Yeh, Luke Miratrix, Rahmatri Mardiko, Yuechen Zhao, Jenny Liu, and Krzysztof Z Gajos. 2013. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2049–2058.

[98] Jie Ren, Viju Raghupathi, and Wullianallur Raghupathi. 2020. Understanding the dimensions of medical crowdfunding: A visual analytics approach. *Journal of Medical Internet Research* 22, 7 (2020), e18813.

[99] Lauren Rhue and Lionel P Robert. 2018. Emotional delivery in pro-social crowdfunding success. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.

[100] David Robins and Jason Holmes. 2008. Aesthetics and credibility in web site design. *Information Processing & Management* 44, 1 (2008), 386–399.

[101] Mario Rojas Q, David Masip, Alexander Todorov, and Jordi Vitria. 2011. Automatic prediction of facial trait judgments: Appearance vs. structural models. *PloS one* 6, 8 (2011), e23323.

[102] Richard Russell, Pawan Sinha, Irving Biederman, and Marissa Nederhouser. 2006. Is pigmentation important for face recognition? Evidence from contrast negation. *Perception* 35, 6 (2006), 749–759.

[103] Johannes Sänger, Norman Hänsch, Brian Glass, Zinaida Benenson, Robert Landwirth, and M Angela Sasse. 2016. Look before you leap: improving the users' ability to detect fraud in electronic marketplaces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3870–3882.

[104] Sharath Sasidharan. 2010. The impact of color and product congruency on user trust in B2C E-Commerce. *ABD Journal* 2 (2010), 1–16.

[105] Etienne Schraven, Elco van Burg, Marco van Gelderen, and Enno Masurel. 2020. Predictions of Crowdfunding Campaign Success: The Influence of First Impressions on Accuracy and Positivity. *Journal of Risk and Financial Management* 13, 12 (2020), 331.

[106] Amazon Web Services. 2016. Machine Learning Image and Video Analysis - Amazon Rekognition - Amazon Web Services. https://aws.amazon.com/rekognition/.

[107] Rotem Shneor and Ziaul Haque Munim. 2019. Reward crowdfunding contribution as planned behaviour: An extended framework. *Journal of Business Research* 103 (2019), 56–70.

[108] ShuiDiChou. 2017. shuidichou. https://www.shuidichou.com/.

[109] Michael Siering, Jascha-Alexander Koch, and Amit V Deokar. 2016. Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts. *Journal of Management Information Systems* 33, 2 (2016), 421–455.

[110] Jeremy Snyder. 2016. Crowdfunding for medical care: ethical issues in an emerging health care funding practice. *Hastings Center Report* 46, 6 (2016), 36–42.

[111] Jeremy Snyder and Valorie A Crooks. 2021. Is there room for privacy in medical crowdfunding? *Journal of Medical Ethics* 47, 12 (2021), e49–e49.

[112] Fitzgerald Steele Jr, David Evans, and Rachel Green. 2009. Is your profile picture worth 1000 words? Photo characteristics associated with personality impression agreement. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 3.

[113] Kok Wei Tan, Bernard Tiddeman, and Ian D Stephen. 2018. Skin texture and colour predict perceived health in Asian faces. *Evolution and Human Behavior* 39, 3 (2018), 320–335.

[114] Anita Tusche, Anne Böckler, Philipp Kanske, Fynn-Mathis Trautwein, and Tania Singer. 2016. Decoding the charitable brain: empathy, perspective taking, and attention shifts differentially predict altruistic giving. *Journal of Neuroscience* 36, 17 (2016), 4719–4732.

[115] Saif Ullah and Yulin Zhou. 2020. Gender, anonymity and team: What determines crowdfunding success on Kickstarter. *Journal of Risk and Financial Management* 13, 4 (2020), 80.

[116] Joost Van De Weijer, Cordelia Schmid, and Jakob Verbeek. 2007. Learning color names from real-world images. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1–8.

[117] Greet Van Hoye and Alan M Saks. 2011. The instrumental-symbolic framework: Organisational image and attractiveness of potential applicants and their companions at a job fair. *Applied Psychology* 60, 2 (2011), 311–335.

[118] Kristina P Vatcheva, MinJae Lee, Joseph B McCormick, and Mohammad H Rahbar. 2016. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale, Calif.)* 6, 2 (2016).

[119] Richard JW Vernon, Clare AM Sutherland, Andrew W Young, and Tom Hartley. 2014. Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences* 111, 32 (2014), E3353–E3361.

[120] Salim S Virani, Alvaro Alonso, Hugo J Aparicio, Emelia J Benjamin, Marcio S Bittencourt, Clifton W Callaway, April P Carson, Alanna M Chamberlain, Susan Cheng, Francesca N Delling, et al. 2021. Heart disease and stroke statistics—2021 update: a report from the American Heart Association. *Circulation* 143, 8 (2021), e254–e743.

[121] Luis Von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. 2008. recaptcha: Human-based character recognition via web security measures. *Science* 321, 5895 (2008), 1465–1468.

[122] Junda Wang, Xupin Zhang, and Jiebo Luo. 2021. How COVID-19 Have Changed Crowdfunding: Evidence From GoFundMe. *ArXiv* abs/2106.09981 (2021).

[123] Xiaopan Wang, Yi Wu, Junpeng Guo, and Lusi Yang. 2020. Says what in your photos? The impacts of photographic narratives on medical crowdfunding performance.. In *PACIS*. 231.

[124] Xingjie Wei and David Stillwell. 2017. How smart does your profile image look? Estimating intelligence from social network profile images. In *Proceedings of the tenth ACM international conference on web search and data mining*. 33–40.

[125] Wang Wei-Ning, Yu Ying-Lin, and Jiang Sheng-Ming. 2006. Image retrieval by emotional semantics: A study of emotional space and feature extraction. In *2006 IEEE International Conference on Systems, Man and Cybernetics*, Vol. 4. IEEE, 3534–3539.

[126] Ziming Wu, Yulun Jiang, Yiding Liu, and Xiaojuan Ma. 2020. Predicting and diagnosing user engagement with mobile ui animation via a data-driven approach. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[127] Ziming Wu, Taewook Kim, Quan Li, and Xiaojuan Ma. 2019. Understanding and modeling user-perceived brand personality from mobile application uis. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[128] Ziming Wu, Zhida Sun, Taewook Kim, Manuele Reani, Caroline Jay, and Xiaojuan Ma. 2018. Mediating color filter exploration with color theme semantics derived from social curation data. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–24.

[129] Yuan Xuan, Yang Chen, Huiying Li, Pan Hui, and Lei Shi. 2016. Lbsnshield: Malicious account detection in location-based social networks. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. 437–440.

[130] Kangning Yang, Chaofan Wang, Zhanna Sarsenbayeva, Benjamin Tag, Tilman Dingler, Greg Wadley, and Jorge Goncalves. 2020. Benchmarking commercial emotion detection systems using realistic distortions of facial image datasets.

The Visual Computer (2020), 1–20.

[131] Elham Yazdani, Anindita Chakravarty, and Jeffrey Inman. 2021. Identifying Emotions in Images and Their Effects on Donation Behavior in Online Crowdsourcing Platforms. *Available at SSRN 3778357* (2021).

[132] Chung-En Yu, Selina Yuqing Xie, and Jun Wen. 2020. Coloring the destination: The role of color psychology on Instagram. *Tourism Management* 80 (2020), 104110.

[133] Marco Zenone and Jeremy Snyder. 2019. Fraud in medical crowdfunding: A typology of publicized cases and policy recommendations. *Policy & Internet* 11, 2 (2019), 215–234.

[134] Fuguo Zhang, Bingyu Xue, Yiran Li, Hui Li, Qihua Liu, et al. 2021. Effect of Textual Features on the Success of Medical Crowdfunding: Model Development and Econometric Analysis from the Tencent Charity Platform. *Journal of Medical Internet Research* 23, 6 (2021), e22395.

[135] Haisu Zhang and Weizhi Chen. 2019. Crowdfunding technological innovations: Interaction between consumer benefits and rewards. *Technovation* 84 (2019), 11–20.

[136] Quanshi Zhang, Yu Yang, Haotian Ma, and Ying Nian Wu. 2019. Interpreting cnns via decision trees. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 6261–6270.

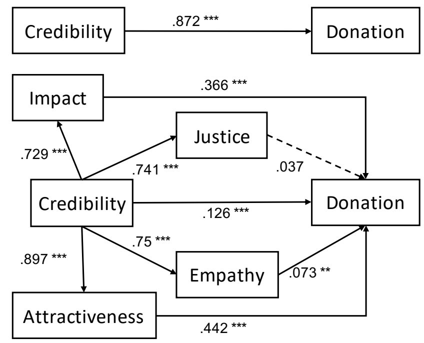
[137] Xing Zhang, Wenli Hu, Quan Xiao, et al. 2021. Influences of Medical Crowdfunding Website Design Features on Trust and Intention to Donate: Controlled Laboratory Experiment. *Journal of medical Internet research* 23, 5 (2021), e25554.

[138] Xupin Zhang, Hanjia Lyu, and Jiebo Luo. 2021. What Contributes to a Crowdfunding Campaign's Success? Evidence and Analyses from GoFundMe Data. *Journal of Social Computing* 2, 2 (2021), 183–192. https://doi.org/10.23919/JSC.2021.0010

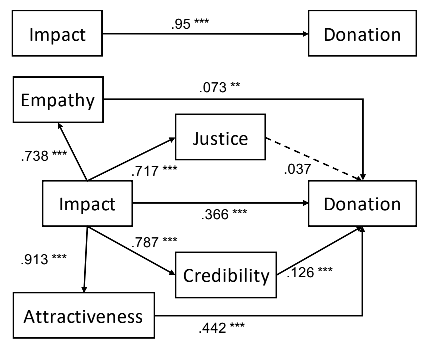
A APPENDIX

A.1 Mediation Analysis for Impressions

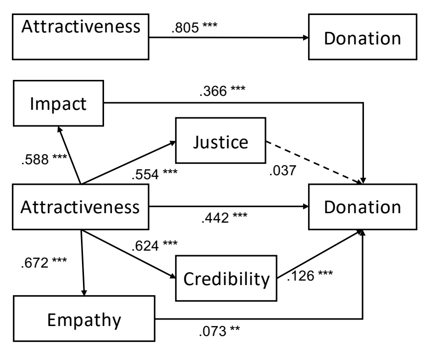
The mediation analysis for perceived first impressions toward donation intention are shown in Figure 10.



(a) Perceived credibility.



(b) Perceived impact.



(c) Perceived attractiveness.

Figure 10: Mediation analysis for perceived first impressions toward donation intention. The solid line stands for $p < 0.05$, and the dash line stands for $p > 0.05$, **: $p < 0.01$, ***: $p < 0.001$.