Feature Article

Challenges in Realizing Smartphone-Based Health Sensing

Alex Mariakakis
University of Washington

Mayank Goel
Carnegie Mellon University

Edward Wang
University of Washington

Shwetak Patel
University of Washington

***Abstract*—This paper presents three case studies involving smartphone-based health sensing projects undertaken by our team. We highlight key challenges that we have encountered while advancing these projects beyond their pilot stages and propose potential directions for engineers, manufacturers, and researchers to address such challenges in the future.**

Advances in science and technology continue to revolutionize medical screening and monitoring. However, people do not benefit from these advances nearly as much as they could. The more frequently a person visits a clinic, the sooner they can receive a diagnosis; once diagnosed, treatment can be delivered more effectively as the patient can monitor their condition more often. However, people typically only schedule yearly checkups, if any at all, because it is not feasible for them to go to the doctor's office more frequently. Even if people are able to make weekly or monthly clinic visits, doing so puts a significant strain on healthcare systems. In the end, reactive healthcare is far more common than preventative healthcare.

People who do not have convenient access to healthcare services are in an even worse position to take advantage of advances in medicine. The World Health Organization estimates that as of 2017, 44% of its member states fall below their recommendation of one physician per 1000 citizens.[1] Even in nations that meet the WHO's recommendation, people sometimes live 3–4 hours away from the closest clinic. Many people in those circumstances rely on community health workers (CHWs) to periodically bring all the

Digital Object Identifier 10.1109/MPRV.2019.2907007
Date of current version 25 July 2019.

1536-1268 © 2019 IEEE
Published by the IEEE Computer Society

necessary screening and monitoring equipment to them; however, CHWs can only carry so many tests with them and must work with a limited budget and training.[2]

We believe that these issues arise because most medical devices are designed with accuracy as their main goal rather than deployability beyond hospitals and clinics. Although accuracy is critical for health applications, there is an opportunity for a new class of medical devices that make small sacrifices in accuracy to enable wider coverage. The mobile health (mHealth) movement aims to support people's health-related needs with existing mobile devices like smartphones and wearables. mHealth can evoke a number of objectives, including the dissemination of medical information and tighter communication between clinicians and their patients. Our research group focuses on the development of smartphone-based health sensing apps. Smartphones include sensors such as accelerometers, microphones, and cameras that can be leveraged to mimic a variety of medical devices. A simple app download could give anyone, anywhere the ability to perform diagnostic screening tests and monitor their condition if or when they receive an official diagnosis.

We have explored various medical subdomains, including hematology,[3–5] spirometry,[6] and cardiology.[7] By making a conscious effort to extend these apps beyond pilot studies, we have uncovered a number of challenges in bringing smartphone-based health sensing to fruition in today's medical and technological infrastructure. These challenges include: (1) limitations fundamental to smartphones themselves, (2) the heterogeneity of smartphone specifications, (3) quality control for measurements in-the-wild, and (4) helping untrained users rationally interpret their data.

In this paper, we share our experiences through a subset of our projects as a series of case studies. We then describe how the aforementioned challenges have manifested in these projects. We discuss how we have addressed some of the challenges in our own work and propose more generalized solutions that can help streamline the realization of these ideas, both in terms of new research directions and manufacturer recommendations.

CASE STUDY #1: BILICAM AND BILISCREEN

Jaundice is the yellowing of the skin and eyes due to the buildup of a compound in the blood stream called bilirubin. Jaundice is a common occurrence in newborns since their livers have not developed sufficiently to break down bilirubin; if left untreated, elevated bilirubin can lead to brain damage. Adults can develop jaundice as well due to conditions that affect the liver and pancreas, such as alcoholism, hepatitis, and pancreatic cancer.

In the past, clinicians have been able to manually analyze photographs of newborns with Photoshop to assess jaundice.[8] BiliCam[4] automates that process using computer vision and machine learning so that anyone with a smartphone can do the same. To use the app, a user takes a picture of the newborn's skin. The app summarizes the skin's color as a feature vector composed of values from multiple color spaces (e.g., RGB and HSV). Those features are mapped to blood test results collected during training to produce a machine learning model that estimates a bilirubin level for future images. Needless to say, skin tone is an important factor when assessing jaundice in the skin, but BiliCam has generalized well given the richness of our training data. In a study with 530 newborns, BiliCam had a mean error of 0.01 ± 1.8 mg/dl and a rank-order correlation of 0.91 when compared to the gold-standard blood draw measurements.[9]

BiliCam does not work for adults because the unhealthy bilirubin level for an adult is an order of magnitude less than that of a newborn (1.5 mg/dl versus 15.0 mg/dl); this level is difficult to detect in the skin, yet more noticeable in the sclera—the white part of the eye. Therefore, we developed BiliScreen[5] as an analogous app for adults. The machine learning pipeline is similar to BiliCam's, but instead of segmenting the user's skin, BiliScreen segments the sclera. The initial 70-person study for BiliScreen yielded a mean error of -0.09 ± 2.8 mg/dl and a Pearson correlation coefficient of up to 0.89 when compared against the blood draw.

CASE STUDY #2: SPIROSMART

Pulmonologists often assess a patient's lung function by having them perform a test through

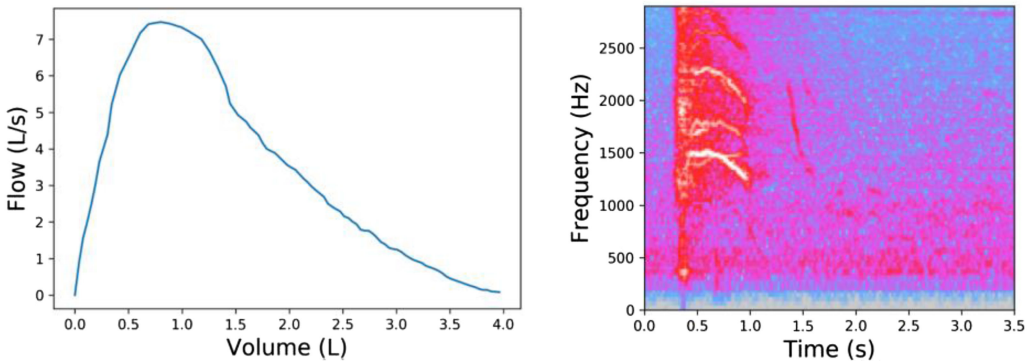

Figure 1. (Left) SpiroSmart uses the smartphone's microphone as an uncalibrated flow sensor to produce flow-versus-volume curves that represents the user's lung function. **(Right)** Features like the resonance of the trachea are extracted from the audio's spectrogram to train SpiroSmart's machine learning algorithm.

spirometry. Lung function is central to the diagnosis of chronic lung conditions such as asthma, chronic obstructive pulmonary disease, and cystic fibrosis. To use a clinical spirometer, the patient inhales to fill their lungs and then exhales through a mouthpiece as forcefully and for as long as possible. The mouthpiece is attached to a sensor that measures the flow rate of the air leaving the patient's lungs, and volume is calculated through cumulative air flow. This data results in a flow-versus-volume curve that can be summarized using metrics such as peak expiratory flow and forced vital capacity. Pulmonologists analyze these metrics and track them over time to monitor a patient's condition.

SpiroSmart[6] uses the smartphone's microphone as an uncalibrated flow sensor for measuring lung function. Microphones measure sound, which is a pressure wave. According to Bernoulli's principle, the flow rate of a fluid like air is inversely related to pressure; therefore, a microphone measures an increase in flow rate as a decrease in pressure. Past research has demonstrated that a microphone can be used to classify wheezing according to flow rate.[10] SpiroSmart uses a similar principle to produce the same flow-versus-volume curves that a spirometer would. To perform the test, a user holds their phone out at roughly arm's-length, inhales, and then exhales as if they were using a spirometer mouthpiece—mouth wide open while forcing as much air as they can for as long as possible. Like a flute, a person's trachea becomes quieter when it is obstructed and higher pitched when it is restricted. SpiroSmart combines both time- and frequency-domain features to infer flow. The time-domain features encode characteristics of the sound's envelope; the frequency-domain features are extracted through resonance tracking on the sound's spectrogram (see Figure 1). Regression models are trained to estimate the flow-versus-volume curves from these features. With models trained on data from 5000 participants, SpiroSmart achieves a mean error rate of 5.7% for FEV1%, the most common measure of lung function; as a point of comparison, the American Thoracic Society requires spirometers to be within 7%–10% of one another on such measures.[11]

CASE STUDY #3: HEMAAPP

Hemoglobin is a protein in red blood cells that carries oxygen through the body. A low hemoglobin level, known as anemia, can often indicate that a person's body is producing less hemoglobin than it is breaking down. Anemia due to malnutrition is one of the biggest reasons behind infant mortality and stunted childhood development in low-resource regions. Anemia is also common amongst pregnant women since their bodies must provide sufficient nutrients to both themselves and their children. The best way to measure hemoglobin is through a blood draw. Recently, devices such as the Masimo Pronto have been developed for measuring hemoglobin noninvasively using a technique called hemachrome analysis. This technique entails shining various wavelengths of light through the fingertip and measuring how much light is reflected or absorbed. As blood rushes in and out of the finger at each heartbeat, only the

absorption due to the blood changes; absorption by the surrounding tissue is constant. The hemoglobin concentration is calculated by comparing the absorption ratios across wavelengths.

Instead of using a dedicated light source and sensor, HemaApp[3,12] relies on the smartphone's existing hardware. Prior work has shown that a person can measure their pulse by covering the smartphone's camera and flash with their finger and having the smartphone measure the change in brightness as the heart beats.[13] HemaApp goes a step further, estimating the hemoglobin concentration against total blood volume by analyzing the color channels. HemaApp uses machine learning to estimate the absorption coefficients of hemoglobin and plasma at the smartphone flash's broadband wavelengths. The features include the intensity of the measured light source, the amplitude of the intensity variation over time, and intensity ratios between the peaks and troughs for each of the different color channels independently. The initial 31-person HemaApp study yielded an RMSE of 1.26 g/dL and a rank-order correlation of 0.82 when compared against a blood draw, both of which are comparable to the Masimo Pronto.

OPEN CHALLENGES IN SMARTPHONE-BASED HEALTH SENSING

Challenge #1: Limitations Fundamental to Smartphones

Most smartphone sensors are primarily focused on improving the user experience. IMUs measure the smartphone's orientation to determine how content should be presented, microphones record audio for communication, and cameras allow users to capture images and videos of their favorite moments. Because these user experiences are currently the primary driving force behind smartphone sales, sensor specifications do not exceed what is necessary to support them.

For example, CMOS image sensors used for smartphone cameras are sensitive to visible and near infrared wavelengths (400–1000 nm). However, most smartphone manufacturers place a thin film on top of the sensor to block infrared light, limiting the spectrum to 400–700 nm to ensure that photographs are visually correct. This design decision for common photography use-cases is counterproductive to specific use-cases such as HemaApp that could benefit from an extended light spectrum. The design of the flash LED also poses challenges for both HemaApp and BiliScreen. The LED is intended for flash photography and torch lighting, so it is designed to produce intense light. For BiliScreen, that intensity can cause discomfort to someone who stares directly at the light. The flash LED and camera also get hot if they are left on for too long, which can cause discomfort while using HemaApp.

Our Approach Although the IR blocking film presents difficulties for HemaApp, some IR light can still leak to the camera if enough is shone. The initial study of HemaApp exploited this fact by utilizing a custom IR and visible light LED array with an incandescent light bulb to augment the smartphone's limited spectral range. The study revealed that incorporating the extra LEDs improved the rank order correlation coefficient between HemaApp's estimates and the corresponding blood draw result from 0.69 to 0.82 when compared to only using the built-in white LED. The use of custom lighting is less attractive than being able to use what already exists on smartphones. Conveniently, newer models have an IR time-of-flight autofocus sensor positioned right next to the rear camera. The current Android API does not provide access to the raw data, but rather the end result of an algorithm that estimates distance. We have accessed this data through a custom kernel installation.

Fortunately, smartphone operating systems have begun to give low-level access to some sensors. In the context of HemaApp, standard white-balancing algorithms often suppress blue and green channel fluctuation because the red channel fluctuation from the blood is so dominant. The Camera 2 API for Android allows for control over such gains, which HemaApp leverages for consistent variation across the color channels. Smartphone operating systems have also begun to provide access to raw image files, which are useful for apps such as BiliCam and BiliScreen that require the truest representation of color directly from the camera sensor.

Figure 2. To account for different lighting conditions and camera sensors, both **(left)** BiliCam and **(right)** BiliScreen incorporate paper accessories with colored squares that can be used as calibration references.

Future Directions A smartphone operating system that offers more control over sensors and other smartphone components can accelerate exploration at the intersection of health and mobile sensing, especially when developers have access to raw sensor data. For example, the IR time-of-flight sensor can be used as a pulse sensor if an API exposes raw sensor values, avoiding the need for custom kernel solutions that cannot be widely deployed.

A loftier goal would be for smartphone manufacturers and researchers to come together and agree upon a concise set of sensors that together form a "dedicated health sensor." Our approach to health sensing has been to push the limits of sensors that already exist on smartphones, yet history has shown that manufacturers are willing to support new sensors if their use has enough value proposition. Apple's M-series coprocessors offload the collection of accelerometer and gyroscope data from the main CPU for gesture recognition even when the smartphone is asleep, and dedicated depth sensors are beginning to appear on newer smartphones for augmented reality applications. Demonstrating the utility of new sensors often requires working with dedicated hardware and then identifying the minimum requirements needed to support the application. This approach can also uncover signals that may not have been discovered otherwise by limiting research to smartphone sensors.

Challenge #2: Smartphone Heterogeneity

HCI and ubiquitous computing researchers often cite the fact that smartphones are pervasive, but this statement only applies to the general category of smartphones; not all smartphones are created equal. There are multiple smartphone manufacturers (e.g., Samsung, Apple, and Motorola) and software operating systems (e.g., Android and iOS), which lead to a diverse smartphone ecosystem. This poses challenges when someone wants to receive FDA approval for an app that relies on the built-in sensors of whichever smartphone model they happened to use for prototyping. The FDA has spent years devising regulations about dedicated medical devices ranging from MRI machines to blood glucose monitors—devices that are assumed to be static and self-contained, performing only their prescribed function with a fixed hardware and software specification.

The studies presented in this paper were conducted using a single smartphone model to avoid cross-device biases. Attaining FDA approval for those apps would require further studies with many different smartphone configurations. Camera-based apps such as HemaApp or BiliScreen, for example, would have to work for a number of different camera modules, LEDs, and sensor arrangements. The flow detected by the microphone in SpiroSmart relies on the mechanical transduction of sound, which is affected by the position of the microphone and the physical casing surrounding it. If generalizability is not possible, developers must restrict potential users to a subset of devices or convince manufacturers to fulfill specific hardware and software requirements to support their app.

Our Approach In BiliCam and BiliScreen, smartphone diversity is handled by performing a check on the camera's properties during data collection. Both apps include paper accessories

for color calibration: a square card for BiliCam and glasses for BiliScreen (see Figure 2). These accessories are inspired by a Macbeth ColorChecker, a professional tool for post-hoc color balancing. If an accessory's colored squares appear different from what was expected, whether due to ambient lighting or the camera's sensitivity to various wavelengths, then the same artifact is likely affecting the appearance of the skin. A calibration matrix that corrects the discrepancy can be applied to the rest of the image to standardize colors across images.

Future Directions Requiring an accessory for standardization adds another potential point of failure that must be FDA-approved. If the BiliCam card's colors fade over time while the card is kept in a person's wallet, the app's performance worsens. The card must also be printed with the same ink and paper used to train the algorithm. In the end, a seemingly trivial addition requires so much consideration that people would probably not be allowed to print the card themselves. Although we posit that such accessories would be far less expensive than a dedicated device, requiring an extra component limits deployability.

Another solution is to create transfer functions based on sensor specifications. When a complete transfer function cannot be generated between sensors, such as two microphones with different sampling rates, compensation mechanisms can be introduced to cater to the common denominator. For developers to find detailed information on a particular sensor, they must currently either disassemble the smartphone and look up the sensor's part number online or dig through the software's kernel and hope the information is documented. Having part numbers accessible in a centralized database or API would help developers understand the capabilities of the sensors at their disposal and account for the diversity in the market. At the minimum, this would allow developers to restrict their app's use to compatible models or software states.

Challenge #3: Quality Control of Data Collection Procedures

Clinical tests are conducted under the supervision of a trained professional. With spirometers, for example, pulmonologists can ensure that their patients use the mouthpiece properly by placing their lips around the tube rather than within it. Pulmonologists can also coach patients on how to properly perform the breathing maneuver so that a spirometer can properly measure their peak and total lung function. Going from using a spirometer in a clinic to using SpiroSmart at home removes that safety blanket of quality control. If a user does not push their lungs to the limit while using SpiroSmart, they can be left with nonsensical results that are not representative of their health. Environmental factors are also more controlled in clinical settings. Traditional spirometers are accurate because they measure flow directly and their mouthpieces block out ambient noise. For SpiroSmart, however, the microphone picks up all the sound that occurs during the measurement, adding unexpected noise to the data.

Enforcing quality control is not only important for the immediate results that people receive, but also for algorithm development. The more assumptions that can be made about the signal, the easier it is for a researcher to design a signal processing pipeline or a machine learning algorithm that arrives at an accurate model. Data collection with many edge cases leads to outliers that either impede system accuracy or need to be handled explicitly.

Our Approach Automatic checks can be implemented to assess the ambient environment before data collection. For example, the BiliCam and BiliScreen apps check that there are no significant shadows or glare spots obstructing the color references. Real-time visualizations can also be made to coach users on how to improve data quality. The SpiroSmart app includes a dynamic visualization that reacts to the flow rate sensed by the microphone in order to encourage users to exhale as much air out of their lungs as possible (see Figure 3, left).

When environmental factors or physical abilities impede a person's ability to comply with data collection, inexpensive accessories can improve the process. One of the observations from the first SpiroSmart deployment was that people with severely impaired lung function sometimes struggled to keep their mouths wide open as they performed the breathing

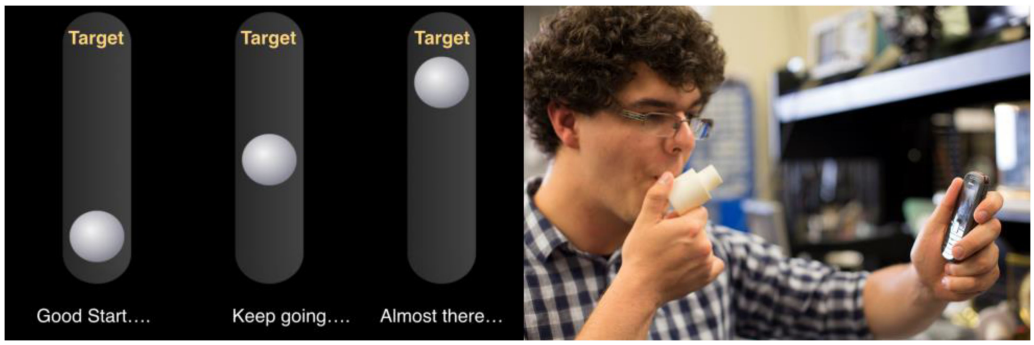

Figure 3. (Left) As a person pushes more air out of their lungs, the ball in the SpiroSmart visualization rises to the top and encourages to the user to continue the maneuver. **(Right)** The SpiroSmart vortex whistle can be used to control the diameter of the user's mouth, acting as a flow-to-pitch transducer.

maneuver. To help those people, a three-dimensional-printable vortex whistle was developed to hold a person's mouth open like the mouthpiece does for a spirometer (see Figure 3, right). The vortex whistle has an extra useful property: the faster the flow of air that enters it, the higher the pitch that leaves it. In other words, the vortex whistle acts as a flow-to-pitch transducer that simplifies the sensing problem.

Future Directions Another way quality control can be integrated into an app is by adding a classifier that decides whether or not data is "valid" before it goes to the main analysis component. In the case of spirometry, researchers had already categorized the mistakes made during spirometry maneuvers (e.g., coughing during the test, pursing lips while blowing).[14] We have trained a machine learning algorithm that identifies these errors for spirometer maneuvers in order to provide users with targeted feedback so that they can improve their technique.[15] We are currently expanding this approach to SpiroSmart, as well.

Challenge #4: Data Interpretation for Untrained Users

The acceleration of hypochondria due to information available on the Internet, also known as cyberchondria,[16] is likely to be exacerbated by ubiquitous medical testing. Using BiliScreen as a worst-case scenario, users could interpret a positive test result as a pancreatic cancer diagnosis. However, not everyone with an elevated bilirubin has pancreatic cancer, and not everyone with pancreatic cancer has an elevated bilirubin. Even if users can internalize this subtlety, false positives and false negatives have significant repercussions, whether it be undue stress or a missed diagnosis.

Doctors receive years of training on how to apply Bayesian reasoning when accounting for a test result in the diagnostic process. This procedure requires calculating the patient's pretest probability of having the condition and then updating that probability according to the accuracy and result of the test. Calculating the prior probability requires knowing the prevalence of the condition and the specific risk factors that may increase a person's likelihood of having the condition, such as family history and environmental factors. Updating to a post-test probability given a positive test result entails calculating the positive predictive value (PPV) of a test—how often people with a positive test result actually have the medical condition. Calculating PPV requires knowing the test's sensitivity (true positive rate), the test's specificity (true negative rate), and the prevalence of the condition:

$$\mathrm{PPV} = \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})}. \quad (1)$$

This calculation does not always lead to intuitive results. A test with a sensitivity and specificity of 80% for a disease that occurs in 15% of the population will have a PPV of 41.3%. A similar

test for a disease that occurs in 5% of the population will have a PPV of only 17.4%. In both cases, the test performs worse than random chance despite having a seemingly decent accuracy.

Test results are never black-and-white; all models have uncertainty bounds that complicate decisions. For example, the current state of BiliScreen has a mean error of -0.09 ± 2.76 mg/dl. This is reasonable for a disease management scenario when a person's bilirubin may vary between 5–20 mg/dl. For a diagnostic scenario, where the threshold for concern is around 1.3 mg/dl, it is debatable whether or not a test result of 2 mg/dl should be considered elevated.

Future Directions We have yet to tackle this challenge ourselves, but we foresee significant contributions that can be made by researchers. If smartphone-based health sensing apps are going to be freely distributed to the general public rather than prescribed and supervised by trained physicians, the routine of estimating a post-test probability should be as automated as possible. Apps should be able to calculate a pretest probability by collecting risk factor information. Family history and demographic data can be explicitly recorded through digital forms. Sensors can also be used to infer risk factors. As an example, GPS data could reveal that a person is at a higher risk of a lung condition because of poor local air quality.

The weighing scale provides an interesting study of how important the presentation of results can be to the decision-making process. All scales have uncertainty, yet people tend to fixate on the number they see. Weight is also a function of how much the person is wearing and how much they ate and drank before the measurement. Kay *et al.* found that people often forget these factors, leading to stress over negligible weight changes.[17] One scale design they suggest graphically emulates a traditional analog scale with exaggerated needle movement to reflect uncertainty. Kay *et al.* also propose an "always-on" scale design that accounts for daily variance and incorporates information through low burden question prompts so that measurements can be automatically adjusted closer to their true value. Researchers in the machine learning community have actually trained models that learn how different clinical measurements vary over time to help clinicians identify high-risk patients.[18,19] The same models could be used to help users extrapolate reasonable trends in their data if they feel the need to do so.

FUTURE OF SMARTPHONE-BASED HEALTH SENSING RESEARCH

Framing smartphone-based health sensing in the context of the existing smartphone infrastructure can be daunting. The smartphones already in people's hands are not designed with health sensing in mind. A reasonable argument can be made that researchers should not restrict themselves to existing hardware from the onset, the alternative approach being to develop custom hardware. One issue we find with this approach is that it limits the potential scalability of a solution. Even if the novel hardware proves to be extremely useful, its uptake will be impeded by the rate at which people purchase the new device, assuming they can afford it in the first place.

We believe that targeting existing smartphones early in the design process allows for more scalable prototypes that can uncover potential issues sooner with a ubiquitous device. Had we not used a smartphone to prototype HemaApp, for example, we would not have discovered that the radiation characteristics of the LED would be integral to the performance. Although today's smartphones may not always be suitable for the solutions developed in research labs, tomorrow's ubiquitous devices can be better informed by those investigations. Hopefully, researchers in this space will address the challenges we have presented in this paper to eventually make medical sensing just a download away.

REFERENCES

1. World Health Organization, WHO | Health workforce, 2017. Available at: http://gamapserver.who.int/gho/interactive_charts/health_workforce/PhysiciansDensity_Total/atlas.html. Accessed on: February 23, 2017.
2. U. Lehmann and D. Sanders, Community Health Workers: What Do We Know About Them? The State of the Evidence on Programmes, Activities, Costs and Impact on Health Outcomes of Using Community Health Workers, 2007.

3. E. J Wang *et al.*, "HemaApp: Noninvasive blood screening of hemoglobin using smartphone cameras," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, vol. 16, pp. 593–604, 2016, doi:10.1145/2971648.2971653.
4. L. de Greef *et al.*, "Bilicam: Using mobile phones to monitor newborn jaundice," in *Proc. UbiComp*, vol. 14, pp. 331–342, 2014.
5. A Mariakakis *et al.*, "BiliScreen: Smartphone-based scleral jaundice monitoring for liver and pancreatic disorders," in *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, pp. 20:1–20:26, 2017.
6. E. C Larson *et al.*, "SpiroSmart: Using a microphone to measure lung function on a mobile phone," in *Proc. UbiComp*, 2012, vol. 12, pp. 280–289, doi:10.1145/2370216.2370261.
7. E. J. Wang *et al.*, "Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera," in *Proc. CHI*, 2018, vol. 18, Paper 425, doi:10.1145/3173574.3173999.
8. S. Leartveravat, "Transcutaneous bilirubin measurement in full term neonate by digital camera," *Med. J. Srisaket Surin Buriram Hosp.*, vol. 24, pp. 105–118, 2009.
9. J. A. Taylor *et al.*, "Use of a smartphone app to assess neonatal jaundice," *Pediatrics*, vol. 140, 2017, Paper e20170312.
10. A. Homs-Corbera, J. A. Fiz, J. Morera, and R. Jané, "Time-frequency detection and analysis of wheezes during forced exhalation," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 1, pp. 182–186, Jan. 2004.
11. A. Rubini, A. Parmagnani, and M. Bondì, "Daily variations in lung volume measurements in young healthy adults," *Biol. Rhythm Res.*, vol. 42, pp. 261–265, 2011.
12. E. J. Wang, W. Li, J. Zhu, R. Rana, and S. N. Patel, "Noninvasive hemoglobin measurement using unmodified smartphone camera and white flash," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2017, pp. EMBS 2333–2336, doi:10.1109/EMBC.2017.8037323.
13. K. Banitsas *et al.*, "A simple algorithm to monitor HR for real time treatment applications," in *Proc. 9th Int. Conf. Inf. Technol. Appl. Biomed.*, vol. 44, pp. 1–5, 2009.
14. L.-A. F. Beeckman-Wagner and D Freeland, "Spirometry quality assurance: Common errors and their impact on test results, 2012.
15. A. Z. Luo, E. Whitmire, J. W. Stout, D. Martenson, and S. Patel, "Automatic characterization of user errors in spirometry," in *Proc. EMBC*, vol. 17, pp. 4239–4242, 2017, doi:10.1109/EMBC.2017.8037792.
16. R. W. White and E. Horvitz, "Cyberchondria: Studies of the escalation of medical concerns in web search," *ACM Trans. Inf. Syst.*, vol. 27, pp. 1–37, 2009.
17. M. Kay, D. Morris, M. Schraefel, and J. A. Kientz, "There's no such thing as gaining a pound," in *Proc. UbiComp*, 2013, vol. 13, pp. 401–410. doi:10.1145/2493432.2493456
18. D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: Using analytics to identify and manage high-risk and high-cost patients," *Health Aff.*, vol. 33, pp. 1123–1131, 2014.
19. S. Saria, D. Koller, and A. Penn, "Learning individual and population level traits from clinical temporal data," in *Proc. Neural Inform. Process. Syst., Predictive Models Pers. Med. Workshop*, 2010, pp. 1–9, doi:10.1016/j.jhevol.2007.05.008.

Alex Mariakakis is currently working toward the Ph.D. degree at the School of Computer Science, the University of Washington, Seattle, WA, USA. His main research is on the creation of mobile apps that use built-in sensors to address diagnostic health and public safety. Contact him at atm15@cs.washington.edu.

Edward Wang is currently working toward Ph.D. degree at the Department of Electrical and Computer Engineering, the University of Washington, Seattle, WA, USA. His main research interest is on enabling health monitoring with nonmedical grade infrastructure through creative sensor usage and signal processing. Contact him at ejaywang@uw.edu.

Mayank Goel is an Assistant Professor with Carnegie Mellon University, Pittsburgh, PA, USA. His research interests include designing, implementing, and testing new sensing systems that can be deployed ubiquitously for health and novel interaction. Contact him at mayank@cs.cmu.edu.

Shwetak Patel is a Professor in computer science and engineering and electrical engineering with the University of Washington, Seattle, WA, USA. His research interests include sensor-enabled embedded systems, HCI, ubicomp, and mobile health. Contact him at shwetak@cs.washington.edu.