Ubiquitous Computing Platforms as Scalable Surrogates in Public Health and Well-being

Joseph Breda

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2026

Reading Committee:

Shwetak N. Patel, Chair

Vikram Iyer

Kurtis Heimerl

Program Authorized to Offer Degree:

Paul G. Allen School of Computer Science & Engineering

University of Washington

Abstract

Ubiquitous Computing Platforms as Scalable Surrogates in Public Health and Well-being

Joseph Breda

Chair of the Supervisory Committee:
Washington Research Foundation Entrepreneurship Endowed Professor Shwetak N. Patel
Paul G Allen School of Computer Science & Engineering

Computing platforms with Sensing components have become ubiquitous across modern society — present in our pockets, on our bodies, and in our environments. While these sensors were originally designed with specific purposes in mind, they can be re-contextualized to offer new insights for novel applications when integrated into controlled studies to couple them with novel signals. Because of their prevalence, these devices become an interface with the population, meaning any findings from studies including them may translate more directly to population-scale. These ubiquitous commodity devices can therefore function as a population-scale development platform for deploying accessible interventions (i.e., diagnostic resources) or extract population-scale insights (i.e., passive sensing).

In this thesis, I explore three approaches to re-contextualizing or extending ubiquitous devices for new sensing applications in public health and societal well-being. First, I demonstrate how the internal temperature sensor of smartphones can be re-contextualized as a thermometer, making fever monitoring more accessible at scale. Next, I show that passively monitored biosignal trends from smartwatches can serve as early indicators of influenza infection. Finally, I show that integrating proximity sensors on bicycles can passively map a surrogate measure of road safety across the network, serving as a means of injury prevention.

Through these approaches, I show how leveraging ubiquitous computing devices can make recovering public health signals more accessible, timely, convenient, and even preemptive.

Contents

Dedicated to Mom, Dad, and Grandpa for teaching me resilience, robustness, and grit.

Acknowledgments

The following (lightly edited) acknowledgments were written in the spring of 2025 shortly before my defense.

I would like to acknowledge that I owe who I am today to the incredible support, kindness, and generosity of an amazing community of friends and mentors that I have been deeply fortunate to end up in. Every achievement I have made in my academic career thus far has been either directly facilitated by, or indirectly due to lessons I've learned from, the people in this community. These same people also made this experience really fun.

Firstly, I would like to thank my committee members who have each played a unique role in my growth as a researcher and as a person.

First and foremost, thank you **Shwetak** for being, not just a strong professional mentor, but a personal role model as well. I learned just as much from your kindness, generosity, and support as a leader as I did from your academic advising. Thank you for taking a chance on me and giving me the space and support to develop into who I am today. Thanks for all the formal advice and all the informal lessons from fun stories over lab lunches and after big presentations over the years. You've created an incredibly positive and unique space within the Allen School, and I feel very luck to have been a part of it.

Thank you **Kurtis** for your encouragement, support, candor, and company. You've always believed in me and given me just what I needed to challenge myself. You've got an exceptional talent for realism, bringing things back to earth, and mitigating self-doubt. You taught me to believe in myself as a researcher, as a community leader, and as a climber at SBP. You also enabled and encouraged me to travel the world, resulting in some of the best experiences of my life. It's been a pleasure traveling with you, running AGENT with you, and learning from you along the way.

Thank you **Vikram** for being such a positive, supportive, and engaging presence in the lab. My PhD wouldn't have been the same if you weren't there in the lab with us every day. You've guided me through hours of brainstorming, soldered with me side-by-side, and brought a new life back into the lab after the pandemic. My PhD experience, and the Ubicomp Lab, truly wouldn't be the same without you.

I would also like to thank my academic mentors outside of my committee and even outside of the university. Thank you **Alex Mariakakis** for your seemingly endless support and generosity. You never owed me a second of your time, and yet never hesitated to step in as soon as I asked. I think the first time I ever asked you for advice, we spent 3 hours on zoom. If that isn't support, I don't know what is. Your advice improved my writing, how I run experiments, and how I deal with uncertainty in my career. Your mentorship early on in my PhD set the stage for all the success I've experienced since. Thank you **Thomas Plötz** for channeling our shared passion for cycling and helping me to turn what I work on into what I love. You were a huge help addressing revisions and taught me a lot about how to frame my work. You've been extremely kind, inspiring, and uplifting ever since the moment we met.

I would also like to acknowledge my friends throughout the department who have made the Allen School feel like a second home. When I've been asked what I got out of the PhD, the first thing that comes to mind is always my incredibly intelligent, supportive, and creative friends. The relationships you make are almost certainly the most important outcome of a PhD and I am particularly proud of mine.

First, I would like to thank **Mike Merrill** for always having my back. You've demonstrated to me time and time again the importance of putting relationships first and what it really means to be a good friend. Thank you **Galen Weld** for consistently going out of your way to brighten my day. You were one of the first people to welcome me to CSE and probably the single most consistent person to visit my desk despite working in an entirely different building.

I would also like to thank all my colleagues and friends who sat with me in the Ubicomp Lab: **Richard Li**, **Shirley Xue**, **Girish Narayanswamy**, **Anandghan Waghmare**, **Ishan Chatterjee**, **Zachary Englhardt**, **Jerry Cao**, **Alex Ching**, **Vidya Srinivas**, **Jiexin Ding**, **Zhihan Zhang**, **Poojita Garg**, **Lisa Elkin**, **Alvin Cao**, **Jason Hoffman**, **Antonio Glenn**, **Matt Whitehill**, **Manuja Sharma**, **Xin Liu**, **CJ Park**, **Farshid Salemi Parizi**, and **Alex Mariakakis**. I feel extremely fortunate to have shared this space with you and been a part of such a friendly, energizing, and innovative community. I'd also like to thank the past members of the Ubicomp Lab who graduated before us. Your success was an inspiration to all of us and set the tone for what the lab is today.

Special thanks to **Richard Li** and **Shirley Xue** for many years of sarcastic banter, Chinese New Year celebrations, hardware advice, and the occasional "hey, hey" or "HELLO!". **Richard**, you taught me many things, but above all, you showed me the importance of just doing things. There are many things I would have never attempted without your excitement and encouragement. A special thanks to **Girish Narayanswamy** for the hours we spent (probably cumulative months of our lives) somewhere between brainstorming and rehearsing comedy routines in the kitchen, and for a surprisingly productive trip to DC. A special thanks to **Anandghan Waghmare** and **Ishan Chatterjee** for some great memories at CHI 2023, exploring Copenhagen together, and for staying up all night in Berlin. **Ishan**, your strategic thinking inspires me and I hope to one day be as good of a party host as you. **Anand**, you've been a great friend and an even better example of how to get things done. You set the pace for everyone in the lab. I'm thankful I got to share so many coffees with you. Finally, a special thanks to **Zachary Englhardt** for your refreshing takes, esoteric domain knowledge, and a fun 2 weeks in Japan min-maxing step count together. I hope to see you in San Francisco after you graduate.

I'd also like to acknowledge my great friends and collaborators both across the department and beyond. A special thanks to **Daniel Campos Zamora** for your grounded and thoughtful advice, and a fun time biking with Zach in Onomichi. I strive to replicate your intuition, foresight, and eye for design. I'm grateful to have you as a collaborator and as a friend. Thank you **Firn Tieanklin** for sharing so many memorable experiences with me in the temples, night markets, and Asian mega-malls of Seoul, Bangkok, Chiang Mai, Kyoto, and now Hanoi. Thank you **Jacob Peplinski** for being my first friend in Seattle, for jumping at the opportunity to spend 5 days on the road to *just go East* together, and for orchestrating a phenomenal lads trip to Mount Baker. Thank you **Dan Sturm** for sharing many years of entertaining and peculiar mannerisms, **Nancy Liu** for always lightening the mood, **Nino Migineishvili** for the down to earth conversations, **Chris Rytting** for many laughs in the year we overlapped, and everyone else who contributed to making this experience a positive one.

I'd also like to thank my mentors in industry **Matt Barnes** and **Keyu Chen**. **Matt**, I appreciated your AI mentorship early on in my PhD in such an apprentice-oriented discipline. It was also great talking urbanism with you. **Keyu**, meeting you was a tipping point in my PhD. Your advice and mentorship gave me a new energy to take on my final projects. I would not have had the confidence to see ProxiCycle through to the end if not for you.

Thank you to my undergraduate advisor **Jay Taneja**, I wouldn't have done any of this if not for you. You helped me publish my first and second research papers, connected me with my future mentors, and encouraged me to pursue a PhD at a time when I still wasn't even sure what I was qualified for. Your mentorship and encouragement literally changed my entire professional career, and for that I am extremely grateful.

And lastly, thank you **Mom** and **Dad**, to whom this dissertation is dedicated. It goes without saying that every success I've ever experienced started with you. **Dad**, you showed me how to be hard working, robust, and convicted, and **Mom**, you showed me how to balance that out with kindness, gratitude, and forgiveness. These complementary qualities have allowed me to be a compassionate and resilient realist and to give everyone a chance, even myself.

The following acknowledgments were appended in the summer of 2026 shortly before handing in this dissertation.

Even though I wrote the majority of this document over a year ago, I couldn't hand it in without adding my thanks to **Daniel McDuff** for being an incredible "post-doc" mentor and friend. You have a really strong taste for research (and sarcasm) and a talent for making things happen which I'm hoping to mirror in my career. It's been a pleasure working with you this year and I'm looking forward to many more to come.

Also, thank you **Jake Sunshine** for your pragmatism and guidance on study design and scientific communication. You have a uniquely valuable perspective and I've already caught myself thinking "how would Jake write this?" Additional thanks to all my collaborators at Google over the last year. I feel incredibly lucky to have landed in such an great community immediately after my PhD.

Thanks **Dean Light** for carrying the torch and making sure CSE is not without an AGENT president or a remarkably sarcastic presence now that Galen and I are supposed to be graduated.

And of course, thanks to all the new members of the ubicomp lab who joined: **Cynthia Dong**, **Chun-Cheng Chang**, **Liam Bakar**, **Guilin Hu**, and **Xiyuan Shen** for keeping the lab energy strong. I hope to see you all in future lab reunions and I am excited to see where your careers take you.

Chapter 1

Introduction and Motivation

1.1 Leveraging Ubiquitous Computing for Societal Impact

Developing scalable solutions for challenges at the population scale typically requires rolling out significant infrastructure. Even measuring simple population metrics like census statistics are massive coordinated efforts. Therefore many insights or interventions at the population scale require some form of inference (i.e., using a more accessible signal to answer a question indirectly) to circumvent the need for costly deployments. The ideal system would even leverage infrastructure that already exists and is deployed at the population scale.

In the modern context, computing and sensing hardware is ubiquitous, already thoroughly deployed at scale across society through smart consumer technology such as smartphones (owned by 91% of US adults [1]), wearables (owned by 1-in-4 US adults [2]), and other embedded systems (e.g., smart speakers owned by 1-in-3 US adults [2]). Because these systems are extensible (i.e., can be easily modified through software and API calls), they form a population-scale development *platform*, presenting an *opportunity* for scaled societal impact. Seizing this opportunity can take the form of leveraging, extending, and re-purposing these existing systems and their components or aspects of their design (e.g., hardware sensors on devices, mobile computing power, form-factors coupled to the body, software traces that measure human behavior through user interaction, etc.) to measure new and useful signals about society at population scale, which can facilitate more accessible, more precise, and sometimes even more preemptive response to users' needs.

While each new generation of consumer technology improves upon the last for a set of tasks

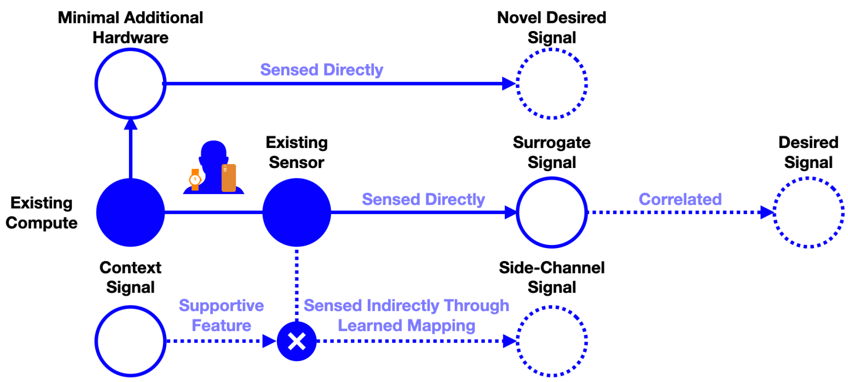

Figure 1.1: An illustration of three approaches to extend existing sensors to capture a novel signal or insight. Surrogate sensing uses the sensor signal directly as a correlate to another signal, while side-channel sensing leverages software methods like machine learning to map the sensor data to a new signal. Lastly, the existing sensor or baseline compute of a device can be further instrumented with minimal hardware to augment or complement its use case.

it was designed to address, my work broadens their scope to novel applications either through secondary (a.k.a. "side-channel") sensing, where a sensor is used to sense a signal other than its designated purpose (e.g., a smartphone accelerometer originally intended to track user activity being used as a communication channel to receive encoded vibrations from a coupled vibration motor), or by treating existing signals captured by these systems as "surrogates", for something useful and correlated (e.g., smartphone usage patterns to measure mental state). These approaches to extend ubiquitous computing and sensing systems are illustrated in Figure 1.1. Either approach extends the capacity of existing sensing infrastructure in a way that can facilitate scaling up a novel insight with little to no added hardware (and in the ideal case can scale to all smart device users with a single software update).

1.1.1 Aligning Population-Scale Technology with Societal Needs

Sociotechnical systems theory asserts that the true value of any technology is unlocked only when it is explicitly aligned with societal objectives [3]. For ubiquitous computing, this may be achieved by treating these thoroughly deployed systems as a population-scale platform for something like capturing population health insights for more rapidly deploying interventions at scale during epidemics (e.g., leveraging location-aware personal devices for contact tracing as a measure to mitigate spread of infectious disease during the Covid-19 Pandemic [4]). This is of course just one of many possible objectives to align these systems with.

To identify impactful population scale challenges to align with the opportunities ubiquitous computing provides, one can start by working backwards from a well established set like the United Nation's 17 Global Sustainable Development Goals (SDG) [5]. Many of these challenges fall outside the scope of what computing can currently address, but among them are two are particularly well aligned with the potential capacity of ubiquitous computing technology for scalable personal sensing and user interaction, which are:

goal 3: to promote good health and well-being

goal 11: to promote sustainable cities and communities

This is because both health and sustainability benefit from precise and continuous personal sensing (of both physiological and environmental factors) and depend on collective action and well informed top-down policy which may be founded on insights aggregated across the population. That is, personal sensing can enable a deeper understanding of one's health, but once aggregated across the population may inform a new understanding of a social determinant of health or a health externality of some environmental factor that may have been previously unknown. There is therefore a cyclical reward to leveraging sensing on personal devices where precision sensing enables individuals to act in ways that can protect both themselves and the people around them (i.e., staying home the moment they get sick to mitigating the spread of infectious disease), and additionally insights aggregated across the population can further inform individual choices via added context or even through new understanding only available through population-scale analysis (i.e., identifying "hotspots" for local disease spread which everyone can avoid). While a purpose-specific tool for something like medical diagnostics might provide higher accuracy or precision to those who can afford access to it (optimizing the impacts on the individual), the additional benefits of population insights and subsequent collective action commonly cited in fields like epidemiology are realized only after sensing (e.g., diagnostics) is accessible to the population broadly.

Building off an already population-scale computing backbone by extending the capacity of devices already in deployment can immediately optimize these systems for collectivism and population response by designing monitoring systems at the population scale directly. Later improvements to the precision and accuracy of these methods can further benefit both the individual *and* the population. This epidemiological model is the most familiar illustration of this cyclical reward, but this can be applied outside infectious disease as well. For example, municipalities must run costly studies to understand the effects of changes to the built environment on things like road safety, physical activity and resident health, and the local economy using arduous methods like visual observation [6]. Similarly to sensing externalities of disease outbreaks, personal devices can capture meaningful signals about how populations respond to these changes in aggregate, enabling more rapid analysis for AB-testing design decisions. Then, if for example, a particular piece of infrastructure is empirically undesirable to interface with (i.e., the data shows most people tend to avoid it), individuals can base their own lifestyle decisions off of this information.

As a step towards closing this loop between individualized personal sensing and population-scale action in health, I have focused a portion of my work on leveraging personal smart devices (e.g., smartphones, smartwatches) for accessibly or even passively detecting immune response to infectious disease. By leveraging devices that are always available or continuously sensing the user's physiology, diagnostic systems built on top of them may allow people to react to the knowledge of illness onset as soon as - or even before - symptoms arise and potentially reduce community spread. I then focus the remainder of my work on extending this paradigm of early-detection beyond disease monitoring to public health more broadly by leveraging mobile personal devices for measuring and mapping useful measures of safety across the road network as a bottom-up approach for mitigating road traffic injury. I approach this by evaluating a smartphone-enabled crowdsourced measure of near-misses as a surrogate for geospatial car-cyclist collision data which is used to identify potentially dangerous road segments.

Just as **early detection** of infectious disease seeks to shift the moment of diagnosis earlier in a disease's progression to curb its spread, **predictive** cyclist injury prevention seeks to identify a measurable signal of cyclist safety capable of revealing the environmental preconditions of traffic injury before harm occurs. The timeline from prediction to early detection and response is illustrated in Figure 1.2. Together, these two threads of my work seek to align ubiquitous sensing with the aims of public health, which the CDC defines as "protecting and improving the health of people and their communities by promoting healthy lifestyles, researching disease and injury prevention, and detecting, preventing and responding to infec-

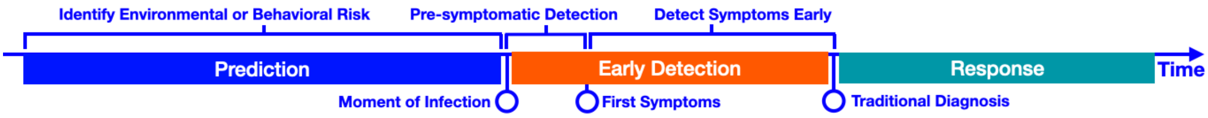

Figure 1.2: The common timeline framework for preemptive public health intervention opportunities for action before a health event (i.e., illness or injury) occurs.

tious diseases"[7][1]. In the following section, I will discuss further motivation for these applications in detail.

1.1.2 Sensing for Infectious Disease Mitigation

Infectious diseases accounted for 14% of global deaths (7.7 million) in 2019 [9]. In the same year 35 million people were sick with the flu and more than half of them made at least one visit to a health care provider [10]. These, often easily preventable, communicable diseases spread throughout communities and consume healthcare resources which has cascading effects on the quality of care throughout the system [11]. This positions early-detection of communicable infectious diseases as highly impactful with a potential to reduce strain on the healthcare system by increasing the individual's ability to self-triage at home as soon as or even before they become contagious [12]. However, existing methods for detecting infectious diseases fail to achieve early-detection as they are often slow and lag behind onset of symptoms, leading to increased likelihood of community spread. For example, the current state-of-the-art method for detecting infection of Covid-19, the flu, and other influenza-like-illness (ILI), is polymerase chain reaction (PCR) tests which require individuals to perform a nasal swab before sending this sample in to a lab for batch testing before finally receiving results. A review of lag time between onset of symptoms and individual testing revealed a median lag of 2 full days (not including the processing time of PCR results and communication delay) [13].

This delay time also varies across regions due to testing resources. For example, a review of PCR result turnaround time in Mexico and Colombia during the Covid-19 pandemic revealed a 3-day median delay between onset of symptoms and test administration for both countries followed by a 7-day median delay between testing and PCR results in Colombia and 4-day me-

[1]This is an adaptation of the original definition of public health proposed by Charles-Edward Amory Winslow in a 1920 Science paper: "the science and art of preventing disease, prolonging life, and promoting health through the organized efforts and informed choices of society, organizations, public and private communities, and individuals."[8]

dian turnaround in Mexico [14]. This delay time leaves individuals undiagnosed while actively contagious which may increase the likelihood of community spread. Additionally, PCR-based testing relies on an individual's intuition that they may be getting sick to decide to take a test which is often after the onset of symptoms and well beyond the point when they become contagious. Their decision to get tested may also be blocked by the cost of testing itself or transportation to a test facility. Ideally, an early-detection system would not just perform as precisely as existing methods but would also do so before infected individuals are even contagious and would be accessible from just about anywhere at no cost.

To this end, a variety of early-detection surrogate signals have been developed at varying scales from classification of individual search engine queries [12] or human activity from fitness trackers [15] to aggregate PCR sampling from urban wastewater[16] or population scale mobility patterns [17]. However, retrospective analysis concludes that the majority of coarse grain estimators such as wastewater analysis would not have curbed the Covid-19 pandemic [18]. Alternatively, simply modeling self-reported symptoms was found to be highly effective at early-detection of covid-19, indicating that sensing specific symptoms or biomarkers directly can have significant impacts on curbing the spread of future pandemics [19]. One way to achieve this is through leveraging ubiquitous technology such as smartphones and smart wearable devices for either continuous passive monitoring or as a platform for accessible self-administered spot-checks to measure symptoms directly and precisely. With as much as 91% of the US adults owning a smartphone [1] and most countries not far behind, these devices serve as the ideal platform for increasing access to diagnostics at a global scale. Additionally, passive biomarker sensing such as signals measured from a smartwatch could identify subtle changes from baselines as a prior probability for infection. The device can then recommend a more accurate self-administered spot-check using a more precise measure from a smartphone. By leveraging ubiquitous commodity devices, this paradigm may scale to all smartphone and smartwatch users, enabling commodity level disease monitoring in near real-time. These individual level measurements could be sent to the cloud and aggregated at the population scale for further analysis. Coupled with existing contact tracking [20] and disease mapping systems [21], this could enable something like interactive participatory epidemiology, closing the loop between individual measurements and population scale insights and policy.

Signals of Infectious Diseases

In this section I will discuss the candidate signals for measuring infectious disease from personal devices. Centers for Disease Control and Prevention (CDC) notes fever (core body temperature of 38 °C or more) as not only a symptom common to all infectious diseases, but often the only sign of infection [22]. Similarly, changes from baseline heart rate variability (HRV) are broadly associated with many forms of infection including sepsis [23], post-stroke infection [24], and infectious respiratory illness like Covid-19 [25]. These signals are continuously captured by most commodity wearables and therefore can serve as the foundation for accessible or even passive flu detection systems.

Core-Body Temperature: Although core-body temperature sensing is not yet available in commodity devices, there do exist multiple temperature sensors integrated throughout most commodity devices for measuring the internal device temperature for monitoring device power and performance. These temperature sensors - known as thermistors - offer an opportunity for side-channel sensing in which effects of external heat from the user's body on the thermistor signal can be disaggregated from other components and then mapped to core-body temperature. Because there are thermistors in almost all smart devices, this approach of side-channel sensing scales across devices more quickly than if new hardware components were to be introduced. The question for this signal is therefore how to extract such a signal from a sensor which is designed for an entirely different application to broaden access to self-administered fever detection (similar to the use of an oral thermometer) to all smartphone users?

Cardiovascular Biomarkers: photoplethysmography (PPG) sensors are increasingly common in commodity fitness tracker devices such as Google Fitbit and the Apple Watch [26]. While initially marketed for applications in exercise and sleep monitoring, the PPG signal collected from these devices has been shown to have promising results for health monitoring in early detection of ILI [25]. As they are worn continuously on-body, they are an ideal candidate for passive and continuous health sensing. Although metrics derived from these sensors have already been shown to correlate to diagnosis of and hospitalization due to ILI, the question of how sensitive these sensors are still remains. For example, can the extracted heart rate and HRV from these sensors predict pre-symptomatically, detect asymptomatic cases, or cases with very low viral load which would be typically overlooked by traditional diagnostic methods [26]?

While there are many other signals which show promise for health monitoring, these two signals offer strong opportunities for impact either through passive monitoring or through the significance of the signal for diagnosing ILI.

1.1.3 Sensing for Injury Prevention

An often overlooked, yet core tenet of public health is to reduce the incidence, severity, and impact of injuries. Sensing for injury prevention therefore aims to prevent personal hardship as well as reduce the stress on healthcare systems by keeping people out of the emergency room through informing policy and environmental design. However, unlike infectious illness which occur regularly and commonly present with biomarkers or symptoms of immune response, the preconditions of serious injury are harder to identify and are usually only addressed long after incidents have occurred, leading to traumatic injury or death.

The CDC tracks the cause of death and reports that 20.2% of all unintended death[2] in the US are caused by motor vehicle or traffic accidents, second only to drug use at 43.69% [27]. This also disproportionately affects people between ages 1-44 meaning drug use and traffic accidents make up the majority of deaths in the US below mean life expectancy. While the US is one of the worst offenders for road traffic fatalities, this severity does extend globally as the World Health Organization (WHO) reports traffic accidents to be the 8th leading cause of death worldwide [28]. The effect of these safety risks also extends well beyond fatalities. For example, of the 10 leading causes of nonfatal emergency room visits in the US across age groups, "struck by motor vehicle" is in the top 5 for all ages and is the second most common for ages 0-24 as well as 65+, second only to "unintentional fall" [3].

This is also a problem that is regrettably getting worse every year. The Organization for Economic Co-operation and Development (OECD) provides data on traffic deaths which shows that deaths due to road accidents have been on a steady rise every year since 2010 [29]. This rise is often attributed to the increasing size of motor vehicles, nicknamed "truck bloat", as manufacturers prioritize the sale of larger truck and SUV sized vehicles which are known to be

[2] due to causes other than medical reasons, suicide, homicide, war, and terrorism
[3] `https://wisqars.cdc.gov/lcnf/`

more dangerous yet now make up 57% of cars on the road [30]. This, coupled with urban design that prioritizes high speeds of motor vehicles over safety has led to a hostile environment for those who choose to travel through other means such as walking, biking, or public transit.

This hostility towards these active transit modes has a secondary impact on public health as active transport (transit modes that incorporate physical activity like walking and biking) is time and time again correlated with a decrease in diabetes, cardiovascular disease (leading cause of death globally, accounting for 33% of all annual fatalities worldwide), and all-cause mortality [31]. In this way, the accessibility of active transport in a region has a direct tie to the health and wellness of the people who live there. Specifically, a recent large-scale global study of physical activity across regions of varying walkability found that more walkable environments are associated with significantly more daily steps across all age, gender, and BMI groups, indicating that areas conducive to active transport have not only a significantly safer road network (i.e., fewer road injuries), but also a healthier population who are less likely to fall victim to cardiovascular disease and metabolic syndrome [32].

The most direct approach to address this takes the form of identifying road designs and specific locations which empirically lead to more frequent collisions and addressing them through changes to the built environment or policy. However, this often doesn't happen fast enough, leading to bottom-up approaches to promote walking and cycling or advocate for change. Either way, evidence of safety concerns in the form of data is essential.

Currently, road safety is measured through historic collisions, meaning assessment of road safety is slow and occurs only after people get hurt. This is even worse for cycling safety where collisions with cars often go unreported [33]. So while safety concerns from fear of collisions prevent cycling adoption, there is little recorded data to make deeper safety assessments. A geospatial signal of cycling safety could both be used from the top down to prioritize safety upgrades to the road network as well as bottom up by directing cyclists on safer routes. The latter is especially important for novice cyclists who are scared to make the initial leap into cycling as a mode of transportation.

Signals of Road Safety

One way in which high resolution safety is measured spatially across the road network for motor vehicle based transport is through the use of collision surrogates which are measurable events which are not true collisions but are highly correlated with real collisions. Some examples are near-misses recorded by static cameras or more inference based approaches leveraging ubiquitous sensing such as hard-braking events recorded by inertial measurement units (IMUs) embedded in car computers or the smartphones of passengers [34]. While similar approaches using IMUs have been attempted for cyclists, a notable limitation is that near-misses between cyclists and cars often do not result in hard-braking or swerving, but rather *close-passes* in which a driver comes within an unsafe distance of cyclists while overtaking [35]. This is a near-miss for a type of collision called a *side-swipe* (where a car strikes the cyclist from the side during a pass) and was found to be the most commonly observed type of collision despite being significantly underreported [36].

In the majority of the US, close-passes are defined as overtaking events in which drivers come within 1 meter of the cyclist. This legally defined safety margin is even larger for roads with higher speed limits in some states. This offers a strong baseline for developing close-passes as a cyclist safety surrogate to better understand the safety of road networks without relying on the previous injury to cyclists. In addition, modern proximity sensors are becoming increasingly accessible and robust. For example, the newly released ST Microelectronics (STM) VL53L8 has millimeter scale resolution, costs only $8, and can sense proximity up to 4 meters in outdoor ambient light conditions [37].

The low cost and size of these proximity sensors make them a logical choice for mobile sensing systems for detecting close-passes from a bicycle. By introducing these sensors to the ends of a bicycle handlebar, the proximity of passing cars can be measured in real-time and the rate of close-passes can be associated with fine-grained location acquired from a rider's smartphone to crowdsource a map of close-pass probability and therefore map safety risk along a road network. I then propose the use of this novel safety signal for routing future cyclists on safer routes to avoid collisions before they happen. While there is potential for other more experimental signals like heart rate as a surrogate for cyclist stress to be used, this work represents a first step into ubiquitously crowdsourced bicycle collision surrogate sensing and explores this first order interaction (i.e., the correlation of dangerous driver maneuvers and empirical cyclist safety).

1.2 Thesis Statement

In this thesis, I propose that **sensors on ubiquitous computing platforms like those already built into smartphones and wearables can be extended or repurposed to improve public health and well-being through early detection or even prediction of illness and injury, mitigating harm at the population scale.** I demonstrate this through exploring the following research questions:

RQ.1 What is a signal which can be collected from sensors onboard commodity smartphones with no additional hardware for detection of infections like ILI at scale and how can it be utilized most effectively?

RQ.2 How sensitive are the biomarker signals sensed by commodity smartwatch fitness trackers to the earliest stages of ILI infection and can these signals be used to detect infection pre-symptomatically or even for asymptomatic cases? And finally;

RQ.3 How can this paradigm of early-detection extend beyond symptoms and biomarkers to detect society level behaviors which pose safety risks, and can these insights be used to mitigate these risks by providing aggregate safety feedback?

These three studies - Chapter 3, 4, and 5 of this document - present unique approaches at mitigating health or safety risk through identifying specific indicators of risk more accessibly and earlier in development to empower individuals (and potentially policy makers once aggregated) to react before any further harm is done. In Chapter 3 I answer **RQ.1** through the development and evaluation of a novel temperature sensing smartphone app - FeverPhone - capable of detecting fever - a common symptom of ILI - and discuss specifics about its utilization. In Chapter 4, I explore **RQ.2** through a flu challenge study in which participants' biomarkers are tracked from the moment of infection with the flu until they are fully recovered to understand how biomarkers respond at the earliest stages of infection. Finally, in Chapter 5, I answer **RQ.3** by developing a novel bike sensor to passively crowdsource a predictive measure of cycling safety for injury prevention. The chapters are arranged in order of progressively more predictive and preventative approaches along the timeline of public health intervention starting with early detection all the way to full prevention. Chapter 3 promotes accessible diagnostics for early detection. Chapter 4 pushes the time of ILI detection earlier in the timeline into the prognostic range (i.e., presymptomatic detection). And finally Chapter 5 looks at predictive measures towards prevention as an intervention.

Chapter 2

Background and Related Work

2.1 Sensing for Public Health

In this section I briefly introduce the space of mobile health before I review the current approaches and related work for sensing ILI and applying the hardware built into commodity smartphones and wearables to this end. I focus specifically on methods of spot-checking or continuously monitoring fever and heart rate as these biomarkers are the primary focus of this work.

2.1.1 Mobile Health Technology

Ubiquitous technology has the capacity to significantly democratize the space of healthcare by providing computational diagnostics technology that is readily available through commodity devices such as smartphones and wearables. The simplest demonstration of this was the implementation of location based contacted tracing to mitigate the spread of Covid-19 [38] or similarly the Flu Near You project in which users crowdsourced flu infection through an online portal [21]. While rudimentary in their signal acquisition, these systems were able to make an impact in slowing disease spread through the ubiquity and access to personal technology for reporting exposure either manually or automatically. However, both of these applications still require PCR results in order to report exposure. There are also many examples of health sensing technology used to pre-screen for certain conditions either actively (by leveraging a specific user interaction) or passively using continuous sensing to check when biomarkers differ from baseline.

2.1.2 Passive Continuous Monitoring

Wearable devices like smartwatches or other smart accessories are often tightly coupled with the user's body and remain on throughout the day and even night providing a platform for continuous monitoring. Smartwatches in particular have been used for a variety of health monitoring including: fitness tracking [39], chronic disease monitoring [40], infectious disease detection [41], and even for applications in mental health and psychology [42, 43]. Many of these applications rely specifically on physical activity tracking and changes in heart rate to infer downstream health tasks. Other research has explored the use of novel wearable form-factors such as smart earrings for acquiring additional biomarkers such as core body temperature which are traditionally challenging from wrist-worn form-factors [44]. There are also examples of data processing techniques which are capable of extracting continuous health metrics from audio and video. One of the earliest examples is Video Magnification which leverages domain expertise in biology to extract heart rate from low magnitude periodic motion from facial video [45]. Following work has been built upon this space to extract PPG wave forms remotely from video using deep learning [46]. Examples in the audio space include both fully passive approaches such as cough detection from ambient audio [47] and opportunistic approaches which leverage acoustic sensing techniques such as frequency modulated continuous wave (FMCW) sonar for passively monitoring respiration rate [48]. While still passive, these approaches get closer to spot-checks. Chapter 4 explores the use of a passive opportunistic sensing to extract and compare biomarkers both before and after flu infection.

2.1.3 Active Spot Checking

Unlike smartwatches or other wearables which remain in contact with the user, smartphone use is more variable throughout the day. However, these devices, when compared to wearables, offer a broader set of sensing hardware and a more interactive user-interface. This makes the smartphone a useful tool for user initiated spot-checks. For example, HemaApp [49] and Seismo [50] are smartphone applications which provide blood pressure and pulse transit time estimates respectively but require a user-initiated test leveraging a specific interaction to make measurements (covering the smartphone flash and camera with a finger in the case of HemaApp and placing the smartphone microphone against the chest in the case of Seismo). Smartphones have also been used in improving the reliability of chemical rapid diagnostic tests [51] or entirely replacing purpose built rapid test readers such as in Glucoscreen [52]. All of these systems leverage the use of a user interaction to narrow the context of a sensor signal to allow for easier detection within a given domain during the execution of the test. Chapter 3 explores the use of this approach to estimate core-body temperature from a smartphone's built-in temperature sensors by creating a context in which temperature change seen by the sensor can be attributed to external factors such as the user's body.

2.1.4 Mobile Health Sensing for Influenza-Like Illness

Mobile devices and wearables are increasingly becoming useful for health monitoring, both at personal scale for preliminary health screening [53] and at population scale for tracking the spread of illness and informing epidemiological models [54]. With respect to ILIs, these smart devices have often been used to record people's sleep, heart rate, and mobility as physiological indicators of ILI symptoms. For example, [55] found that heart rate variability differed for the seven days before and after a COVID-19 diagnosis with noticeable changes in the signal observed the first day of reported symptoms. [56] leveraged data on sleep, heart rate, and mobility from smartwatches and fitness trackers to predict cases of COVID-19. Finally, [57] used anonymous location history collected from smartphones to forecast the spread of the flu. Despite the utility of core body temperature as a signal for influenza-like illness and the existence of temperature sensors on-device, these works and others have yet to incorporate it into their mobile health sensing systems.

2.1.5 Body Temperature Sensing with Smart Devices

Because fever is both common to all ILI and oftentimes the sole indicator of infection, it is a primary focus of this work [22]. In the following sections I discuss the existing technology for fever detection and its current limitations as well as advancement in sensing which have implications in fever detection.

Limitations of Consumer-grade Thermometers

Consumer-grade peripheral thermometers come in multiple form factors that can measure temperature at different parts of the body. These locations include but are not limited to the ear canal, the mouth, the armpit, and the temporal artery. Many purpose-built at-home thermometers show a high degree of error, demonstrating that thermometry is still a challenging task. In a review of 75 different studies on the accuracy of peripheral thermometers, [58] found that the studied thermometers had pooled limits of agreement outside the predefined clinically acceptable range of ± 0.5 °C, with errors being as large as ± 1.46 °C for febrile adults. These findings have been corroborated by [59] and [60], among others. Similarly, [61] performed a meta-analysis of consumer-grade infrared temporal thermometers for fever cutoff classification and found that of nine different studies these devices anywhere from 0.41 to 0.91 sensitivity with 0.85 to 1.00 specificity. Another study by [62] found the AUC for non-contact infrared thermometers to be between 0.8 and 0.9 when screening at the forehead site, but also found this value to fall below 0.8 for colder ambient air temperatures (14 °C–16 °C). Nevertheless, consumer-grade peripheral thermometers are currently the standard for at-home temperature monitoring and are used for screening by many businesses.

Emerging Alternative Thermometer Designs

Researchers have explored alternative hardware designs to perform fever detection. For example, dedicated torso- and wrist-mounted sensors have been developed to capture the wearer's skin temperature [63, 64]; this data can later be used to infer core body temperature by accounting for basic information like the wearer's height and weight. Researchers have also explored the use of contactless thermal cameras for fever detection. [65] and [66] leveraged forward-looking infrared (FLIR) cameras to identify fevers from infrared video data. The Huawei Honor Play 4 smartphone released in 2020 includes an infrared temperature sensor which can be used as a thermometer [67], but this sensor has not seen pervasive use in other smartphone models. On the whole, these approaches to temperature sensing require dedicated or specialized hardware that is not yet ubiquitous, limiting the ability for researchers and clinicians to acquire temperature data during telehealth scenarios for personal screening. Our work seeks to alleviate this issue by taking advantage of devices that are already widely adopted, namely smartphones, for this purpose.

2.1.6 Temperature Sensing Using Smartphone Thermistors

Smartphones are equipped with many temperature-sensitive resistors called thermistors for monitoring the integrity of the device's other internal components, namely the battery. This technology is nearly the same hardware present in state-of-the-art clinical-grade thermometer probes. Peripheral thermometers that leverage thermistors often do not measure temperature directly but rather predict body temperature based on the change in temperature experienced by the thermistor when it comes into contact with the body. On top of this temporal prediction, proprietary algorithms are used to map peripheral body temperature to core body temperature. Although smartphone thermistors are designed to monitor the temperature of components in the device itself, they have been repurposed for two main applications, which are listed below.

Ambient Air Temperature

Past work has shown that it is possible to passively measure the ambient air temperature for the purposes of crowdsourcing temperature estimates within buildings or cities [68, 69]. [70] demonstrated ambient air temperature estimation by combining battery temperature data with information about the smartphone's software usage and physical context. This work leveraged a hierarchy of linear and exponential models to characterize the different thermal states of the device, and then it used activity recognition to select the appropriate model. [71] later developed a similar technique for ambient air temperature estimation, utilizing battery current usage to infer the state of the device. Both of these studies specifically modeled the heat generated by smartphone activity and decoupled this information from ambient air temperature. Finally, [72] focused on opportunistically estimating ambient air temperature when the smartphone had minimal software utilization and, therefore, minimal heat being generated within itself. The authors go on to describe how the position of the device (e.g., within the pocket, in the user's hand) impacts the temperature estimates.

Core Body Temperature

The notion of using smartphones to estimate core body temperature has been explored to a limited degree in prior work. [73] ran a simulated study using a sous-vide machine and exhibited the feasibility of using smartphone thermistors to discriminate high and low water temperatures after making contact with the simulated heat source for 3 minutes. [74] replicated this research with similar results after 8 minutes of contact between the device and simulated heat source. Neither of these studies evaluated the corresponding systems for measuring core body temperature in live human subjects. FeverPhone represents the first example of a system that leverages smartphone thermistors for core body temperature estimation and fever monitoring in real people. Prior work related to ambient air temperature estimation has focused on passive sensing scenarios, requiring information about the software being run on the phone and how the phone is being operated by the user to apply corrections to the data. Prior work also entirely focuses on the use of the battery thermistor which is the only publicly accessible thermistor on Android. FeverPhone makes use of all the thermistors on the smartphone by leveraging Android phones with root-level access to read each thermistor's system files. These thermistors were found to update at a higher frequency and perform better at estimating core body temperature than the battery thermistor. FeverPhone is not a passive system but rather relies on an active interaction to estimate core body temperature. This means that both the software running on the device and the way the device is held during the interaction can be controlled to improve accuracy.

2.2 Sensing for Injury Prevention

In this section I discuss the existing body of work on traditional collision surrogates used in injury prevention as well as the space of crowdsourced urban informatics.

2.2.1 Incident Detection

It is essential to track traffic incidents (such as collisions, breakdowns, or other disruptions) to maintain safety of the road network. This is evident in the many systems or agencies carrying this out. For example, in the US alone, the National Highway Traffic Safety Administration (NHTSA)'s Fatality Analysis Reporting System (FARS) [75] or the Federal Highway Administration (FHA)'s Highway Safety Information System (HSIS) [76] conduct surveys and aggregate police reports to serve as a census for car collisions. However, these databases are arduously compiled manually, slow to update, and are inherently sparse signals, which fail to represent rapidly changing environments. For this reason, traffic engineering and urban planning leverages additional data to capture these effects.

2.2.2 Traditional Collision Surrogates

Traditional road safety literature addresses the scarcity of historic collision data through collision surrogates which are measurable events which correlate highly with or even predict collisions. Some examples of existing collision surrogates include *Near Misses*–when vehicles nearly but do not collide– which can be extracted from traffic camera footage [77] or self-reported on crowdsourcing platforms like UpRide [78] as well as *hard-braking events*–when a vehicle stops abruptly– which can be detected by accelerometers either built into or coupled to cars (i.e., smartphone mounted on the dashboard) [34]. Similar paradigms have been used to explore bicycle swerving in response to driver behavior as a potential surrogate such as in CycleSense. However, the authors note that incident types such as tailgating or close passes may go undetected by motion based models as cyclists may not change their motion profile due to these dangerous events. Bicycle motion is notably prone to false positives due to motion artifacts such as braking and riding on uneven terrain as indicated by CycleSense's reported precision of 0.035 [35, 79]. Similarly, Apple Watch's accelerometer based crash detection systems [80] experienced high false positive rates due to the indirectness of the surrogate signal. For example, in January 2023, Apple's crash detection system triggered 134 false emergency calls in Hida Japan's mountain region (more than 14% of all emergency dispatch calls) reportedly due to iPhone 14s triggering while users were skiing [81]. These false positives may be mitigated by deploying more accurate sensing systems which either: (1) include context signals to identify situations which can trigger false positives; or (2) leverage sensors which measure events which are more directly associated with the incident being detected.

2.2.3 Crowdsourcing and Passively Sensing City Scale Informatics

Prior work has addressed the need for scaling data to population levels with crowdsourcing or passive sensing to map the build environment similar to how cyclists build up a mental-map of safety through experience and exploration [82]. However, by leveraging sensing or crowdsourcing, these maps can be aggregated across the population and shared with new users who would previously need to gather this information themselves. For example, Project Sidewalk combines crowdsourced labels and existing Google Street View (GSV) images to identify sidewalk accessibility issues to facilitate better navigation and awareness around sidewalk accessibility hazards [83, 84, 85]. Crowdsourcing alongside GSV has been used for other mapping tasks such as mapping street level objects [86] or even perceived beauty of the built environment [87] to improve city scale monitoring and navigation. Alternative signals such as geolocated social media posts have been used to map urban well-being [88]. Prior work has also demonstrated the feasibility of deploying passive sensing for sourcing entirely new datasets such as city-wide air quality by mounting sensors on vehicles like garbage trucks or passenger cars [89, 90, 91]. Similarly, passively monitored GPS data from smartphones has been used to naturalistically identify disease spread hot spots or influences of the built environment on people's health [92, 93].

While little work has been done to explore technology for quantifying cycling safety through other means, there are many examples of quantitative data improving safety and access to active transport. For example, OneBusAway demonstrated the impact of increased public transit ridership with the introduction of better arrival time predictions and real-time bus location updates [94]. Interestingly, participants of the study cited increased perception of safety traveling by bus as a primary factor influencing their decision to ride more frequently. Later a study of Project Sidewalk, a crowdsourcing system for identifying hazards limiting access on urban sidewalks, found that simply providing quantitative data about the built environment encouraged community members to be more conscious of these issues in their daily lives and impacting their future decision making [85]. Similarly, BikeNet argues that environmental sensing from a bike has the potential to improve cyclist's quality of life by tracking exposure to pollution from cars or informing their regulation of activity during a route using quantitative feedback provided after completing a ride [95].

Chapter 3

Feverphone: Accessible Core-Body Temperature Sensing Using Commodity Smartphones

In this chapter, we discuss the development of a side-channel sensing system to estimate the core-body temperature of a user for fever detection during a brief interaction where the user brings the smartphone screen in contact with their forehead, illustrated in Figure 3.1. This work aims to improve the timeliness of diagnosing of infectious disease at scale by leveraging the existing hardware on smartphones to provide an accessible approach for thermometry. We demonstrate the performance of this system in both an in-lab validation study and a real-world clinical validation. This chapter was published in ACM IMWUT 2023 and presented at Ubicomp 2023 [96] where it won the Distinguished Paper Award at Ubicomp 2024 and was later highlighted to the broader SIGMOBILE community in ACM GetMobile 2025.

3.1 Introduction

Traditionally defined as a core body temperature of greater than 38 °C (100.4 °F), fever is known to be the primary predictive symptom for many viral infections and is the single most commonly cited manifestation of COVID-19 [97, 98, 99, 100, 101, 102]. In many cases, clini-

Figure 3.1: A demonstration of the physical interaction used to test an individual's core body temperature by bringing the smartphone screen in contact with the user's forehead. This interaction allows for the smartphone screen to measure the amount of contact which influences heat exchange which is measured by the device's internal thermistors.

cians use a lower threshold of 37.5 °C (99.5 °F) [97, 103] or even 37.3 °C [104] as the cutoff for a low-grade fever, triggering the early detection of illness and longitudinal monitoring to defend against community spread. A host of thermometer models are available for at-home patient use, ranging from in-ear to skin-surface to oral form factors, but these devices are often not available in people's homes. A study by [105] found that out of a cohort of 141 parents who required frequent temperature monitoring for their child with febrile convulsions, only 21 parents (15%) actually had access to a thermometer at home. The lacking ubiquity of thermometers may be due to some combination of their non-trivial cost ($15–$300 USD) or the fact that they are only typically needed a few times per year. Thermometers can also sell out during times of need, further reducing access [106].

More readily available thermometry with comparable error could have a substantial impact on personal health screening, identifying nascent cases of illness and enabling more informative telehealth consultations. Accessible fever monitoring could also facilitate pandemic mitigation and epidemiological modeling [107]. To achieve these capabilities, we propose a method of estimating core body temperature using commodity smartphones. Smartphones have become increasingly ubiquitous in recent years, with nearly 3 billion smartphones in 2019 [108]; many of these users reside in under-served areas due to their limited proximity to clinical resources or economic constraints. Previous work has explored the use of smartphone-connected thermometers to detect influenza-like illnesses at scale [109], hinting at the potential benefits that a purely smartphone-based solution could have.

In this paper, we leverage the thermistors embedded in smartphones to estimate a person's core body temperature. We do this by collecting thermistor and touchscreen readings over a 90-second interaction where the smartphone touchscreen is held against the user's forehead. We then define a data-driven model that maps features describing the rate of heat transfer between the user's forehead and the device to the ground-truth temperature collected by an oral thermometer. We deployed this technique in a smartphone app called FeverPhone, and we evaluated this app both in lab and in a clinical setting. First, we used a simulated heat source in the form of a temperature-controlled bag of water to ensure the temperature of a source in contact with the device can be recovered from the thermistor signal. Beyond evaluating the core working principle of FeverPhone, we evaluated its robustness across multiple confounds: phone model, phone accessories, pressure of device application, and ambient temperature. We then deployed FeverPhone in a clinical trial with 37 patients, 16 of whom had a low-grade or standard fever. We found that FeverPhone achieves a mean absolute error of 0.229 °C, a limit of agreement of ±0.731 °C, and a Pearson's correlation coefficient of 0.763.

In summary, our primary contribution is a demonstration of how the thermistors in a smartphone can be used to estimate a person's core body temperature without any hardware modifications to the device itself. We demonstrate this capability through the following components:

- A smartphone application and user interaction capable of estimating core body temperature using just the built-in device hardware,
- A lab validation study using a simulated heat source to validate the working principle of FeverPhone and many potential confounds against its generalizability, and
- A preliminary clinical study with 37 patients in a local emergency room.

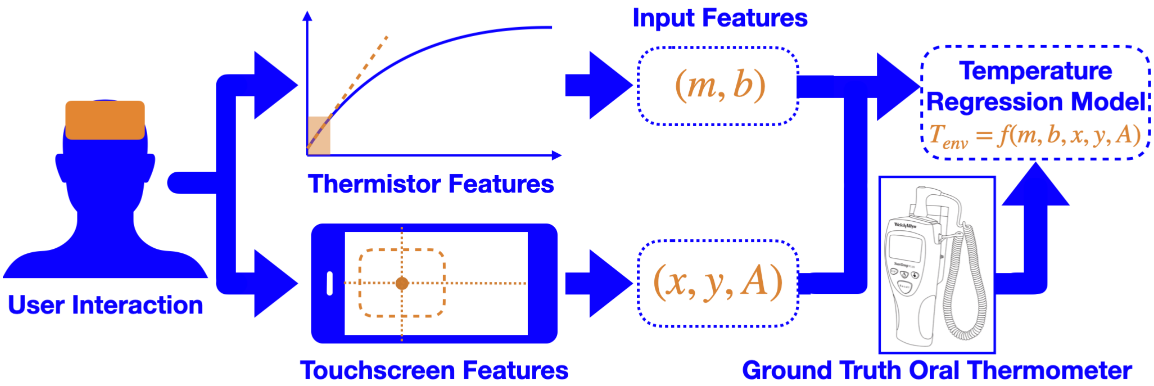

Figure 3.2: A box diagram illustrating the algorithmic pipeline for FeverPhone. As the user places the smartphone's touchscreen against their head, FeverPhone records features related to the temperature change experienced by the thermistor and the contact surface as measured by the phone's touchscreen. These features are used in a regression model to predict core body temperature.

3.2 Sensing Temperature with a Commodity Smartphone

Fig. 3.2 illustrates the overall design of FeverPhone. In short, we treat the smartphone and its built-in thermistors as a temperature probe. The user is asked to place the phone's touchscreen in direct contact with their forehead in order to control the data collection process in a repeatable and convenient manner. FeverPhone then analyzes not only the temperature signal measured by a single thermistor, but also the capacitive data from the device's touchscreen to capture features summarizing the type of contact made with the user's skin. In the subsections that follow, we describe the design decisions and algorithm behind FeverPhone in greater detail.

3.2.1 Working Principle

Newton's Law of Cooling

Newton's Law of Cooling states that the rate of heat loss of an object is directly proportional to the difference between the temperatures of the object and its environment. This concept is mathematically expressed as follows:

$$T(t) = (T_0 - T_{env})e^{-kt} + T_{env} \tag{3.1}$$

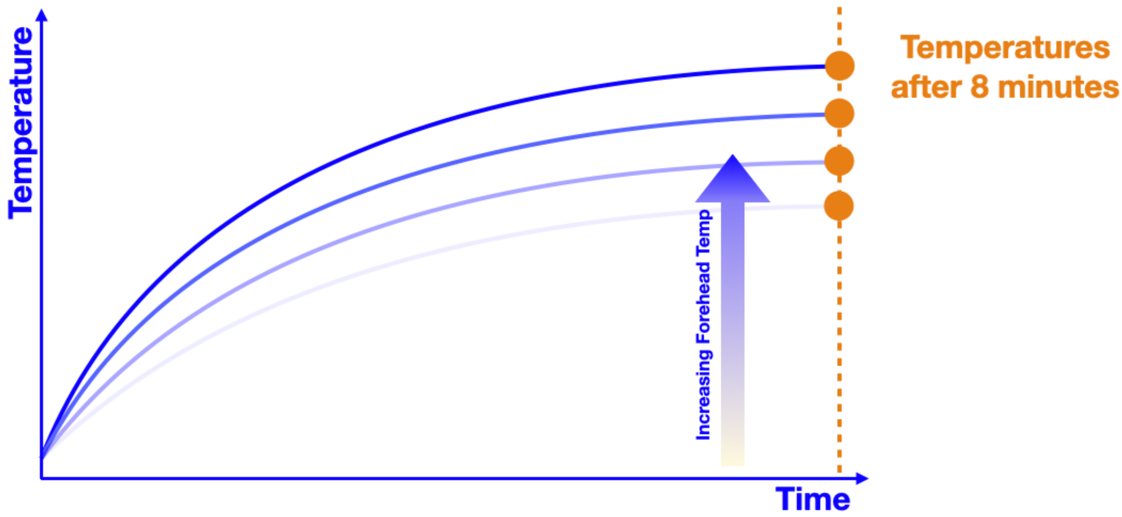

Figure 3.3: An illustration of the increasing steady state temperature reached with Newton's Law of Cooling model at increasing core body temperatures.

where T_0 represents the initial temperature of the object, T_{env} is the temperature of the environment, and k is a constant that dictates the cooling rate of the object.

For the purposes of FeverPhone, T_0 represents the initial temperature measured by a given thermistor before contact is made between the device and the user's forehead. Since the thermal mass of the user is significantly greater than that of the smartphone, the temperature of the device should eventually reach T_{env} — the steady-state temperature reached by the device after it comes to thermal equilibrium with the user's forehead. k describes the rate of heat transfer from the smartphone to the user's forehead, which varies across devices and users. More specifically, device-specific factors include the total mass of the smartphone, the position of the thermistor within the device, and the device's software utilization. User-specific factors include the physical composition of the user and the way that they position the smartphone on their forehead. Although k encompasses many factors, Equation 3.1 is still idealized as it does not consider the cooling of the smartphone that occurs separately from the heat transfer to the user's forehead (e.g., heat dissipation, heat transfer to the user's hand). This is illustrated in Figure 3.3.

Preconditions of Thermometry

Much like existing clinical-grade thermometers, FeverPhone relies on heat being transferred from the user to the thermistor. If the temperature of the thermistor is too similar to the temperature of the user's forehead, there will not be enough heat transfer to make a proper estimate. Further, if the temperature of the thermistor is greater than that of the user, heat will instead be transferred from the device to the user. All of these factors are influenced by the ambient air temperature, which impacts both the resting temperature of the thermistor and the user's peripheral body temperature. Therefore, most thermometers and temperature sensing devices constrain the ambient air temperatures that are suitable for operation. This range is largely based on the thermistor's material[1], and using thermometers outside of these bounds can lead to inaccuracies [110, 111].

As with all other thermometers, FeverPhone requires that the user's body temperature is greater than the temperature of the device. We also require users to take a conservative break of 10 minutes between successive readings so that the thermal probe can reacclimate to its original temperature. We only test FeverPhone when the ambient air is between 18.3 °C and 23.6 °C, which is tighter than the operating ranges of clinical-grade thermometers like the Welch-Allyn SureTemp 690 thermometer[2] (10 °C–40 °C) and consumer-grade digital thermometers like those by Kerma Medical[3] (18 °C–28 °C). Establishing the operating range of FeverPhone and expanding it further would just be a matter of collecting data at more extreme temperatures since FeverPhone is data-driven, so we leave these questions to future work. In Section 3.2.7, we explore the impact of testing FeverPhone when the ambient air temperature is outside what is available during training.

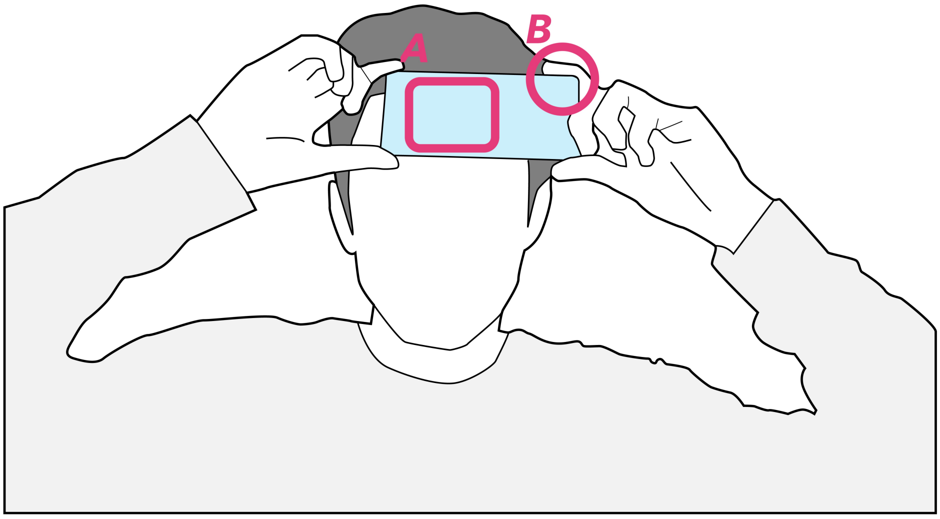

Figure 3.4: Example posture of user interaction. Box A highlights the region of contact between the user's forehead and device screen closer to the top half of the device, and Box B highlights the "camera-grip" style where the four corners of the device are pinched between the user's fingers.

3.2.2 Testing Procedure

Testing Location and Smartphone Positioning

There are various body locations that could be suitable for smartphone thermometry, namely the hands, ears, forehead, and armpits. These testing sites have unique advantages in terms of their ergonomics and affordances, but we have specifically designed FeverPhone with the forehead in mind for two reasons. First, the forehead is less susceptible to confounds like ambient air temperature [112, 113] and physiological thermoregulation [114, 115, 116]. For example, the hands can vary by as much as 15 °C when exposed to varying ambient air temperatures [112], whereas the forehead only varies by 3 °C across a similar range [113]. Second, the forehead has a sufficiently large and exposed surface to make contact with the device's screen. Although the armpit can result in even more heat transfer, the resulting signal can be less consistent due to variation in clothing and body curvature.

We define a standardized interaction to control the user's grip, the orientation of the device, and the region-of-contact between the user and the device as best as possible. By holding

[1] `https://www.electronics-tutorials.ws/io/thermistors.html`

[2] `https://www.welchallyn.com/content/welchallyn/americas/en/products/categories/thermometry/oral-axillary-rectal-thermometers/suretemp-plus-690.html`

[3] `https://www.kermamedical.com/Products/Details/7`

the phone in the "camera-grip" style illustrated in Fig. 3.4, the user minimizes the amount of heat that is transferred from their hand to the device. Holding the smartphone horizontally against the forehead ensures that the contact area spans most of the device's minor axis, and we found during our preliminary investigations that most smartphone thermistors are more sensitive to heat sources in contact with the middle or top-third of the smartphone. We therefore instruct users to prioritize making contact with this region of the device. Sensor data is recorded for a fixed duration that we investigate in our analyses, but the upper limit for the duration of heat transfer is 120 seconds.

Preconditions of FeverPhone

The working principle of thermometry is applicable to FeverPhone as long as the preconditions of thermometry are met and other heat-generating processes are either minimized or properly characterized. Prior work has measured the rate at which various software operations generate heat within smartphones (e.g., high CPU utilization, Wi-Fi or cellular activity) [117, 70]; in some cases, so much heat is generated that the smartphone becomes warmer than the user and the working principle of FeverPhone cannot be applied. In our own experiments, we found that unlocking the screen or plugging in the charger each increase the temperature by 0.5 °C to 1 °C when the device is otherwise idle and unplugged at room temperature (18.3 °C–22.2 °C). We also found that holding the device in hand also raised the temperature by a similar increment.

To minimize these confounds, we only collect data when the smartphone is in a trickle-charge state with minimal CPU utilization, which requires the device to be fully charged, plugged into an AC-portable charger, in airplane mode, and with the screen brightness fixed to 20%. Most of these constraints can be applied automatically via software and have been imposed in past work related to opportunistic temperature sensing with smartphones [72, 70, 73]. Given that FeverPhone calls for the user to place their smartphone against their forehead, we argue that these restrictions are even more reasonable since people would not be precluded from using their device any more than they would be from the active interaction required by FeverPhone in the first place. We also require the user to hold the smartphone in their hand for 60 seconds before taking a measurement since that is usually the amount of time it takes for

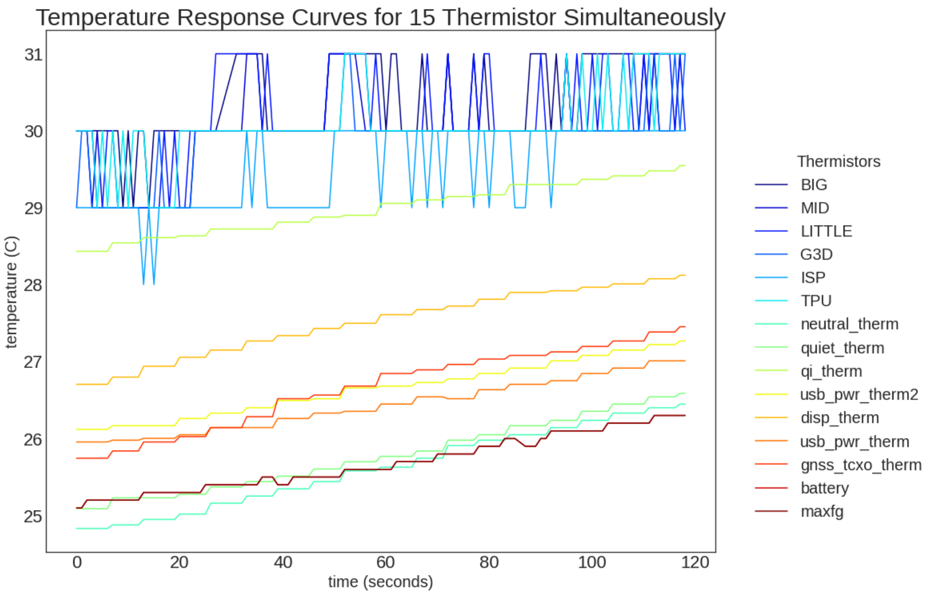

Figure 3.5: An example of the differing thermal response curves for a selection of thermistors on the device during 120 seconds of contact with a human forehead.

the device to reach a steady-state temperature.

3.2.3 Sensor Data Collection and Modeling

Thermistor Data

Various thermistors with unique thermal response curves are distributed throughout different smartphone models. FeverPhone relies upon the thermistor that is most sensitive to heat conduction from an external source in contact with the phone's touchscreen. Fig. 3.5 illustrates the difference in thermal response curves for a selection of thermistors on a Google Pixel 6 while the device was in contact with a human forehead. We found that the strongest signals were acquired from thermistors that had lower resting temperatures. These thermistors exhibited greater dynamic ranges in temperature, and according to Newton's Law of Cooling, steeper rates of heat transfer. The fact that lower initial temperatures lead to better thermal response curves corroborates the preconditions for existing thermometers outlined in Section 3.2.1.

Given a sufficient number of temperature readings over a long period of time, Equation 3.1 could simply be fit to the thermistor data using an exponential model and T_{env} would directly reveal the temperature of the user's forehead. However, getting reliable results using this

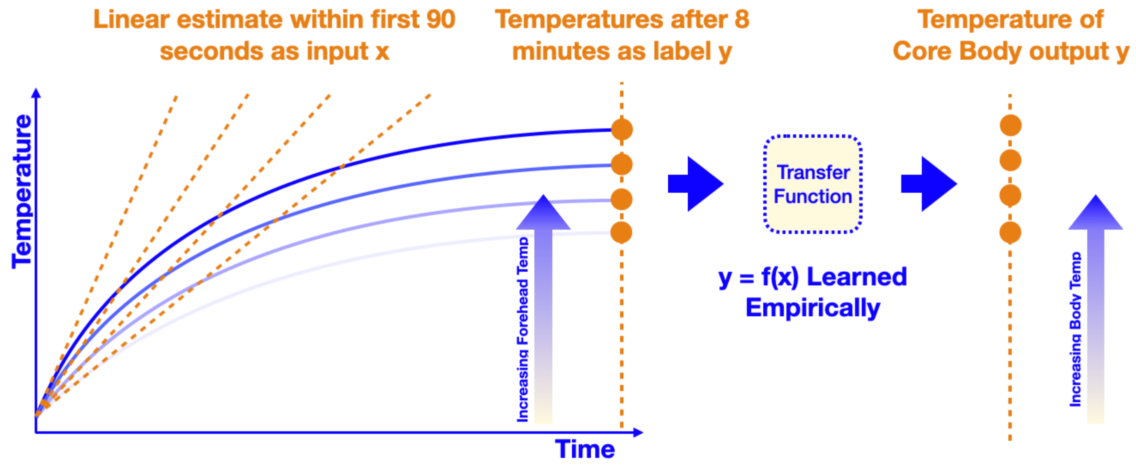

Figure 3.6: An illustration of how the linear approximation within the first 90 seconds of data collection increases to reflect an increased core body temperature and is used for core body temperature prediction.

approach requires longer contact periods (as much as 8 minutes in our experience) since exponential functions are highly sensitive to noise at smaller values of t. To make temperature estimation faster, we rely on the fact that the rate of change in thermistor temperature is relatively consistent during the initial few minutes of the heat transfer since there is a relatively small difference in the initial temperatures of the device and the user's forehead. This allows us to approximate Equation 3.1 as a linear function:

$$T(t) = mt + T_0 \tag{3.2}$$

where m encodes similar characteristics to k from Equation 3.1 along with some approximation of $(T_0 - T_{env})$ — the difference in device and body temperature during contact. We fit the temperature data from the thermistor to this model to calculate the rate-of-change m and the initial temperature T_0 as features in our model. Figure 3.6 illustrates how the linear approximation of Equation 3.1 is used to predict steady state temperature.

Capacitive Touchscreen Data

One of the major factors that influences the thermistor signal observed by FeverPhone is the region of contact between the smartphone and the user's forehead. The closer the region is to the internal location of a given thermistor, the quicker the rate of heat transfer the thermistor experiences. This distance cannot be easily calculated since thermistor locations are typically not public knowledge, but we can account for it by extracting features that summarize where the user has placed their forehead along the smartphone according to the capacitive touchscreen.

Android offers a `TouchEvent` API for extracting the location of touchscreen interactions; however, this API assumes that all interactions are caused by fingers and can therefore be summarized as ellipses. FeverPhone instead uses a user-debug build of Android to read the raw capacitance values from the 2D array of capacitors inside the touchscreen every 5 seconds. We summarize this data by first averaging the capacitive values across all of the recorded matrices. We then mask out the cells with an average capacitance greater than 25% of the maximum capacitance value in the averaged matrix. This produces a binary mask that describes the region of contact created by the user.

To account for the distance between the region of contact and the thermistor, we calculate the vertical position of the region's weighted centroid relative to the phone's edge. To account for the amount of contact between the user's forehead and the device, we calculate the total area of the unmasked region. We explored other features like the maximum capacitance as a proxy for the amount of contact being applied against the smartphone; however, we did not find any correlation between this feature and rate of thermistor temperature change. We suspect that this is because as long as a sufficient amount of pressure is being applied to the smartphone, the rate of heat transfer from the heat source to the smartphone remains the same. However, this could also be a result of the max capacitance not being a perfect proxy of pressure since capacitance values have a limited range.

Machine Learning Model

We train a machine learning regressor to learn the complex interactions between our features as they relate to core body temperature. Our model uses four features for temperature estimation: (1) the initial temperature of the thermistor before contact with the user, (2) the rate of change of the thermistor's temperature over the data collection, (3) the vertical position of the contact surface's weighted centroid, and (4) the total area of the contact surface. As illustrated in Fig. 3.7, we conceptually split the features into two different groups. The rate of change in the thermistor's temperature is considered the primary feature since it is the only

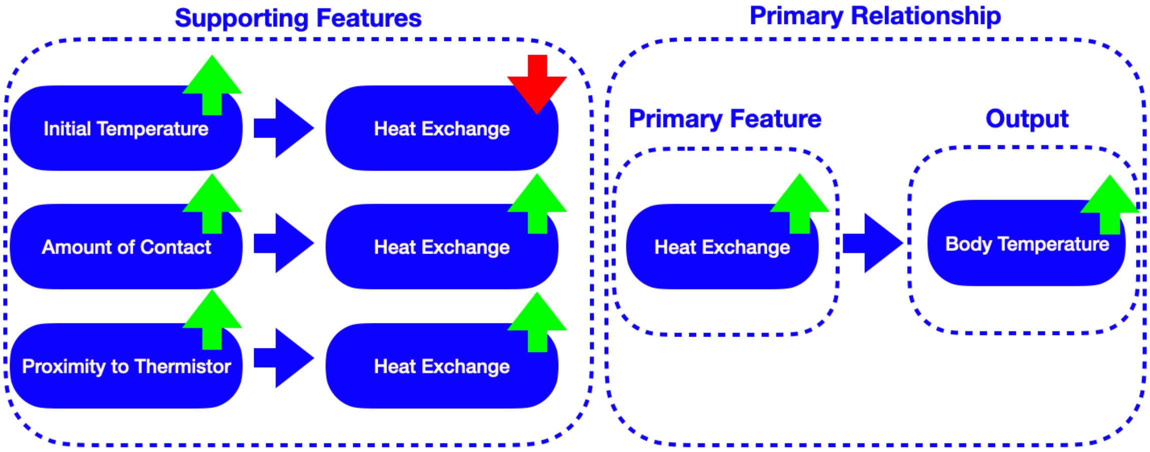

Figure 3.7: A block diagram illustrating the relationship between the features used in FeverPhone. The rate of change in thermistor temperature, or "heat exchange", is the primary variable related to the model output, while the other features provide context for that observation and how it is mapped to the output.

feature that varies with respect to the user's body temperature. The other features are considered to be supporting features, provide context to the thermal curve that is observed by the thermistor.

3.2.4 Study 1: In Lab Validation

Before deploying FeverPhone to human participants, we first validated the working principle of the system in a lab setting using a simulated skin-like heat source: a sous-vide precision water heater. In this experiment we investigate the relationship between features. This validation methodology has been leveraged in prior work for body temperature monitoring [73, 74].

3.2.5 Experimental Procedure

We collected data on a Google Pixel 6 smartphone without any accessories (e.g., phone case, screen protector). We simulated a human forehead using a resealable plastic bag full of water and a PrimoEats sous-vide precision water heater[4]. The experimental setup for the experiment is illustrated in Fig. 3.8. The sous-vide was fixed to the same location inside the plastic bag throughout the experiment. The physical interaction of holding the smartphone against the forehead was emulated by a cardboard clasp that applied enough pressure to the back of

[4]https://www.amazon.com/Cooker-Immersion-Digital-Display-Stainless/dp/B08227CF83

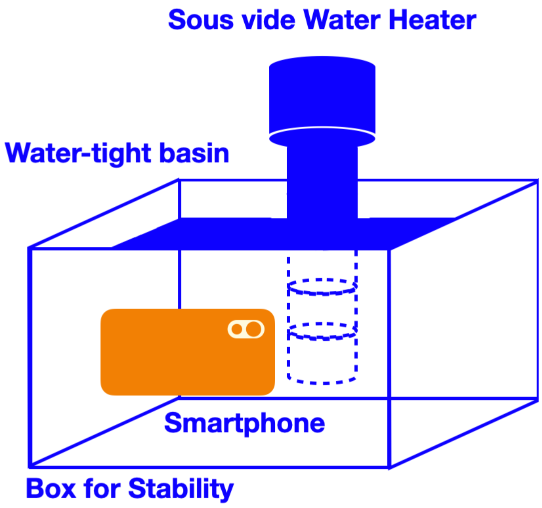

Figure 3.8: An illustration of the experimental setup for the lab validation study. The sous-vide was placed inside a plastic bag filled with water and supported by a cardboard box. The phone was mounted against the cardboard box with a forehead-sized hole cut into the box to allow direct contact between the temperature-controlled bag of water and the phone screen.

the device to make steady contact between the touchscreen and the bag of water. The bag was placed inside a cardboard box with a hole about one-third of the length of the device cut out to allow only a portion of the screen to maintain direct contact with the simulated heat source. This was done to control the region of the touchscreen that made contact with the plastic bag. The study was conducted in a laboratory room with an ambient temperature at roughly 23.6 °C.

We intentionally introduced a degree of variation during the data collection process to reduce the impact of some potential confounds. The smartphone's position was slightly adjusted between trials so that the device made contact with the bag of water at different points along the screen. We also varied the initial temperature of the smartphone by running compute-intensive tasks on the device so that it could heat up until it had reached a target temperature. The initial temperature of the smartphone was varied between 21.1 °C and 28.3 °C. Meanwhile, the sous-vide was used to sustain the water at temperatures between 36.4 °C and 39.2 °C in increments of approximately 0.25 °C[5]. This procedure resulted in the collection of 41 data samples across these dimensions.

[5]The sous-vide model used for this experiment only supported temperatures in Fahrenheit, so the exact temperature was varied in increments of 0.5 °F. The manufacturer of the sous-vide reports that the precision of their device is ±1%, although the baseline temperature from which this variance was reported is unclear.

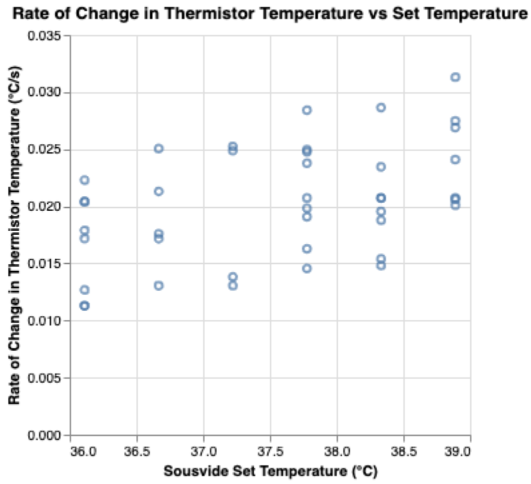

Figure 3.9: A scatter plot showing the linear relationship between the rate of change of the thermistor's temperature and the sous-vide's temperature.

3.2.6 Feature Relationships

As mentioned earlier, the change in the thermistor's temperature is the only feature in FeverPhone that shares a direct relationship with the temperature of the heat source being measured. Fig. 3.9 shows this relationship across all trials of our sous-vide experiment. The Pearson's correlation coefficient (r) between the two variables was 0.44 ($p < .01$), indicating that there is a weak linear correlation between them. To further explore the linear relationship between the rate of change in the thermistor's temperature and the temperature of the sous-vide, we trained single-feature linear regression models using 5-fold cross-validation. This resulted in a mean absolute error (MAE) of 0.786 °C and a goodness of fit (R^2) of 0.194. Given that the temperature of the sous-vide ranged by approximately 3 °C, these results show that the rate of change in the thermistor's temperature provides some useful information for estimating the temperature of the heat source but not enough to be used on its own.

Although there is no meaningful relationship between the temperature of the heat source and either the initial temperature of the phone or the touch-based features, these features do share a direct relationship with the rate of heat transfer to the thermistor. We illustrate these relationships in Figure 3.10, which shows the correlation between each of the supporting features and the rate of change in the thermistor's temperature. The location of contact led to the highest correlation out of the three supporting features, indicating that the heat transferred to the phone from the forehead is not transferred equally across the device and that the prox-

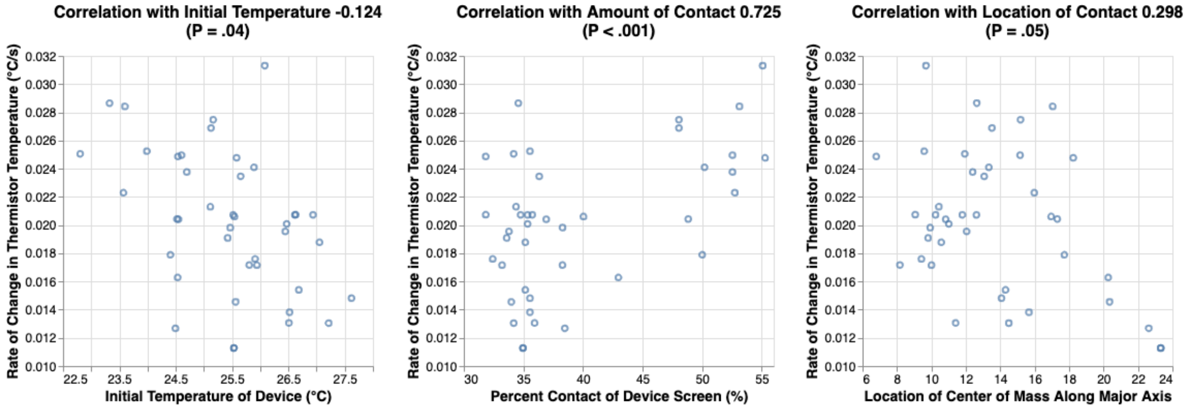

Figure 3.10: Scatter plots showing the relationship between each of FeverPhone's supporting features — (from left to right) initial temperature of the smartphone, the amount of contact as measured by the touchscreen, and the location of the contact along the vertical axis — versus the rate of change of the thermistor's temperature.

imity of contact to the thermistor is very important in FeverPhone. The least correlated of the three supporting features was the initial temperature. Despite the software restrictions we impose on the smartphone during data collection (e.g., airplane mode, fully charged), the smartphone is constantly warming up or cooling down. Therefore, the initial temperature only captures a brief snapshot of the device's state before data collection. We also trained linear regression models that estimate the change in the thermistor's temperature as a function of three supporting features to further support their utility in providing context to the observed rate of heat transfer. As before, the models were trained using 5-fold cross-validation, but we express the error using mean absolute percent error since the relative changes in the output are more meaningful for FeverPhone's final temperature estimate. This experiment resulted in a mean absolute percent error of 9%, a Pearson's r of 0.853, and an R^2 of 0.850. We then trained linear regression models to estimate the sous-vide's temperature as a function of the primary feature and the three supporting features. After training these models using 5-fold cross validation, we achieved an MAE of 0.495 °C, a Pearson's r of 0.864, and an R^2 of 0.715. Adding the supporting features resulted in a 43% decrease in error from using the primary feature alone, showing that the supporting features add important contextual information to the rate of change in thermistor's temperature.

3.2.7 Potential Confounds

There are a number of potential confounders that could influence FeverPhone's generalizability to real-world scenarios. To measure the impact of each one in isolation, we repeated the data collection protocol described in Section 3.2.5 while varying one of the following parameters: the model and instance of the smartphone, the accessories on the smartphone, the pressure applied to the smartphone by the sous-vide bag, or the ambient air temperature. After collecting 4 samples at 3 sous-vide temperatures (36.7 °C, 37.8 °C, and 38.9 °C), we evaluated FeverPhone's sensitivity to these changes in two ways. The first method, which we consider to represent FeverPhone's *uncalibrated* performance, involved training a single model on the original dataset and then evaluating its performance on the new dataset with the confound introduced. The second method, which we consider to represent FeverPhone's *calibrated* performance, involved merging 10-15% (6-10 samples) of the data with confounds present with the original dataset before conducting 5-fold cross-validation; the characteristics of the datasets that were merged varied slightly depending on the nature of the confound being tested, so we clarify these details below before describing the corresponding results. We aim to show that only a small amount of data affected by confounds is necessary for the *calibrated* error to outperform both the *uncalibrated* error and dummy regression without significantly impacting the *uncalibrated* baseline error. Since this evaluation method produced predictions for all data including the baseline dataset, the reported error metrics were stratified according to confound state. The results of these experiments are presented in Table 3.1.

Phone Model & Instance

To test how well FeverPhone generalizes to new devices beyond the Google Pixel 6 used for baseline training, we collected additional data using a second Google Pixel 6, a Google Pixel 3, and a Huawei P20. In the uncalibrated approach, the MAE increased to 0.709 °C when the baseline model was applied on the other Pixel 6. Although the increase in error may indicate that there were subtle differences in the underlying hardware components that were sourced for these two instances of the same device, the moderate change in error indicates that the model was still able to generalize across instances with reasonable accuracy. As expected, the MAE rose significantly when the model trained on the Pixel 6 was deployed on other phone

Confound	Variation	Uncalibrated Test MAE	Uncalibrated Test Pearson's r	Calibrated Test MAE (baseline MAE)	Calibrated Test Pearson's r (Baseline Pearson's r)
Baseline	Dummy regression	1.072 °C	-1.000	—	—
Baseline	Baseline dataset	0.495 °C	0.864	—	—
Phone Model	Different Google Pixel 6	0.709 °C	0.491	0.790 °C (0.680 °C)	0.701 (0.720)
Phone Model	Google Pixel 3	8.76 °C	0.375	0.825 °C (0.852 °C)	0.312 (0.448)
Phone Model	Huawei P20	4.12 °C	0.410	0.649 °C (0.728 °C)	0.599 (0.632)
Accessory	Screen protector	1.094 °C	0.823	0.634 °C (0.664 °C)	0.691 (0.754)
Accessory	Phone case	1.285 °C	0.751	0.692 °C (0.664 °C)	0.678 (0.754)
Accessory	Phone case and screen protector	0.813 °C	0.775	0.695 °C (0.664 °C)	0.797 (0.754)
Pressure	Gentle	0.845 °C	0.878	0.545 °C (0.595 °C)	0.831 (0.804)
Pressure	Forceful	1.606 °C	0.553	0.874 °C (0.595 °C)	0.758 (0.804)
Ambient Temperature	21.3 °C	0.610 °C	0.745	0.584 °C (0.545 °C)	0.766 (0.786)
Ambient Temperature	16.9 °C	1.180 °C	0.952	0.629 °C (0.545 °C)	0.745 (0.786)

Table 3.1: A table of regression metrics for the experiments used to evaluate the impact of various confounds. The baseline model was trained on data collected in a 23.6 °C room using a Google Pixel 6 without any accessories that was being pressed against the sous-vide bag with moderate pressure. Each row indicates the only variable that was modified for each given experiment. For each calibrated result, similar metrics are provided for both the calibration test set and the original baseline test set.

models with significantly different hardware composition (e.g., thermistor location, case material, heat capacity), to the point where they both exceeded the MAE of the dummy regression. Given the heterogeneity across devices, our calibration approach entailed merging the baseline dataset with only the dataset of the new phone being introduced rather than merging all phones into a single dataset before cross-validation. This procedure significantly improved the correlation for the other Google Pixel 6 at the cost of increasing the MAE slightly. Meanwhile, this procedure significantly decreased the MAE for the Huawei P20 and the Google Pixel 3 to be under 1 °C. After calibrating the models for each phone model, the performance those models had on the baseline dataset slightly decreased. These results indicate that FeverPhone can generalize fairly well across different devices with some amount of calibration data, but the benefit is far more profound when the devices vary significantly in hardware.

Phone Accessories

To test FeverPhone's robustness when a person places a phone case or screen protector on their device, we collected additional data with these phone accessories independently and together. In the uncalibrated approach, we found that both accessories increased the MAE near that of the dummy regression but with significantly higher correlation. The error was higher with the phone case than it was with the screen protector, which is likely due to its higher heat capacity. To our surprise, using the two accessories in tandem actually led to better estimates in terms of MAE, although the MAE of that model was still worse than the MAE on the baseline dataset. Since these accessories vary the heat capacity in a continuous manner, our calibration approach entailed merging the baseline dataset with the data for all three variations before cross-validation. This procedure improved the MAE of all three configurations while decreasing their correlation in two cases. The calibrated model also had similar performance on the baseline dataset. These findings indicate that larger or more insulating phone accessories will result in a greater impact on model performance after calibration. Although it is possible to account for these errors with calibration, temporarily removing these accessories may be a more robust way of improving FeverPhone estimates.

Pressure of Device Application

To test how well FeverPhone generalizes to different amounts of pressure being applied by the user, we collected additional data with notably "gentle" and "forceful" pressure. We note that more rigorous definitions of pressure are difficult to specify since the water bag is not rigid and applies a counter-pressure of its own; a load cell would have also absorbed some heat on its own, thereby impacting our measurements. In the uncalibrated approach, the MAE for the gentle and forceful pressure scenarios was 0.845 °C and 1.606 °C respectively. The temperature estimates for the forceful pressure scenario tended to be higher than expected since more heat was being transferred from the sous-vide bag to the thermistor, and the converse was true for the gentle pressure scenario. The significantly worse accuracy in the forceful pressure scenario may be partially attributed to deformation of the sous-vide bag, which may have exposed the smartphone to heat in different ways than anticipated. Since pressure is a continuous quantity, our calibration approach entailed merging the baseline dataset with the data for both of the other pressure applications. The MAE across both of the unseen pressure levels improved after calibration, but the error in the forceful pressure scenario was still relatively high. These findings indicate that FeverPhone is reasonably robust to pressure of application when trained with measurements across different amounts of pressure. Furthermore, forceful pressure may have more of a deleterious impact than gentle pressure.

Ambient Air Temperature

To test how well FeverPhone generalizes to different ambient air temperatures beyond the 23.6 °C room that was used for baseline training, we collected additional data in a 21.3 °C room and a 16.9 °C room. In the uncalibrated approach, the MAE degraded as the ambient air temperature deviated further from the temperature in which the model was trained. This result can be attributed to the fact that the baseline model was never trained on data samples where the initial thermistor temperature reached these colder temperatures and was thus unable to generalize. Since ambient temperature is a continuous quantity, our calibration approach entailed merging the baseline dataset with the data for both of the other room temperatures before cross-validation. Although the MAE at the original room temperature increased by nearly 0.05 °C, the MAE across the colder room temperatures improved to within 0.1 °C of the calibrated MAE on the baseline dataset. These findings indicate that FeverPhone can reasonably generalize to new ambient air temperatures as long as the training dataset includes variety in the initial temperature of the thermistor.

Summary

We observed that all the examined confounds negatively impacted the performance of FeverPhone when the uncalibrated approach was used for testing. Attempting to generalize FeverPhone across completely different phone models, using a phone case, or applying forceful pressure increased the error to worse than the dummy regression, albeit with significantly better Pearson's r. For most of the other confound variations, the MAE remained close to within 1 °C. By incorporating some additional training data across a broader set of conditions (e.g., ambient temperatures, amounts of pressure), we were able to calibrate the model to accommodate these scenarios. The MAE fell within 1 °C for all confound variations without significantly hampering the performance on the baseline dataset.

3.2.8 Study 2: Clinical Validation with Real Patients

Having supported the validity of FeverPhone's design through our lab validation study, we deployed our app to a local emergency department to assess its efficacy with human participants as both a thermometer and a low-grade fever detector. We address **RQ1** through a series of sub-research questions related to the design of FeverPhone along the way:

RQ1.1 How does the selection of the machine learning model impact the performance of FeverPhone?

RQ1.2 How does the duration of data collection impact the performance of FeverPhone?

RQ1.3 How does the inclusion or exclusion of features impact performance of FeverPhone?

RQ1.4 How can FeverPhone detect fevers in the absence of labeled febrile training samples?

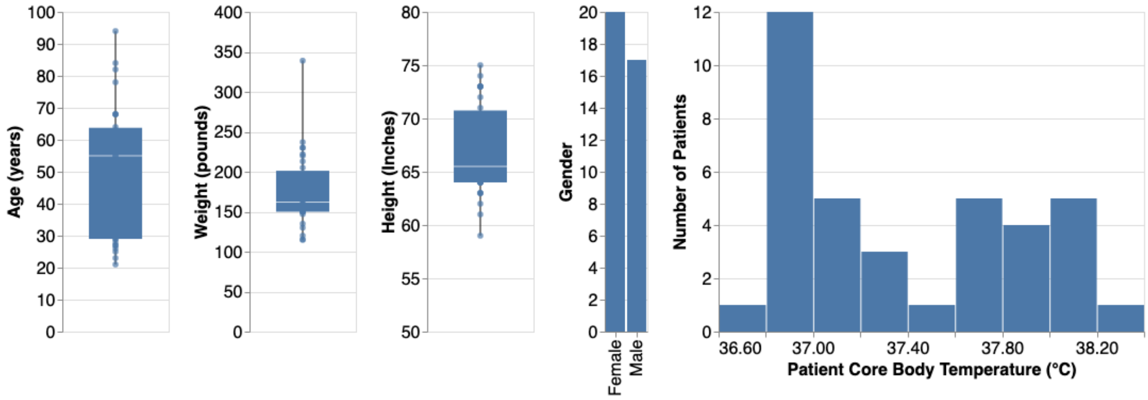

Figure 3.11: The demographic distribution of participants in the clinical study.

3.2.9 Participants

Individuals who arrived at the University of Washington's Emergency Department were given the opportunity to participate in our study regardless of their reason for visiting the hospital. The only inclusion criteria related to the study were that the participants had to be 18 years or older, able to provide informed consent, and in a stable medical condition. Participants were excluded from our protocol if they had their blood drawn immediately before recruitment or had taken an antipyretic drug like Tylenol since both of these actions can rapidly impact core body temperature. A summarization of participant demographics can be found in Fig 3.11.

Prior to enrollment, eligible participants had their temperature measured using a Welch-Allyn SureTemp Plus 690 oral thermometer. This thermometer also provided our ground-truth measurements since it had a reported accuracy of ± 0.1 °C and was able to deliver readings noninvasively within 10 seconds; the convenience of the thermometer was particularly important for minimizing interference with patient care. Participants with nominal core body temperatures ($\sim$37 °C) were often patients who were visiting the hospital for ailments like physical trauma that do not elevate core body temperature. Participants with at least a low-grade fever ($\geq$37.5 °C) were verified by clinicians to have known cause of fever, often a viral or bacterial infection. Participants rarely presented with a high-grade fever above 38.6 °C, and in these cases, their skin was so sweaty that it was difficult to gather a reliable measurement with FeverPhone. This is a known limitation not only with FeverPhone, but also other temporal and peripheral thermometers; temporal thermometers and other skin contact-based thermometers suffer from reduced accuracy when the skin is sweaty or even clammy [118, 119, 120]. Since it is easy to identify fevers in such individuals without thermometry, we excluded these participants from our analyses. In total, we recruited 37 participants (17 male, 20 female), 16 of whom presented with at least a low-grade fever. The average age of our participants was 51 $\pm$ 20 years.

3.2.10 Experimental Procedure

After obtaining informed consent and a ground-truth measurement of the participant's temperature, a research team member measured the participant's temperature using FeverPhone implemented on a Google Pixel 6 smartphone. To avoid interfering with patient care, all trials were conducted opportunistically. Data collection occurred in different parts of the emergency department. Patients in the waiting room and hallway were sitting upright while patients in private rooms were laying down in beds. Since some individuals were too weak to hold the smartphone on their own, the researcher held the smartphone against the participant's forehead for them.

Participants were asked to have their temperature recorded by FeverPhone up to three times, but this was not always possible without impacting patient care. Participants who were willing and able to contribute repeated trials waited 10 minutes in between to allow both their forehead and the smartphone to acclimate back to their steady-state temperatures. The phone was wiped down using a sanitary alcohol wipe between trials, which simultaneously minimized the spread of germs while cooling the device back down between trials. New ground-truth measurements were also collected immediately before each repeated trial.

3.2.11 Analysis

We trained our regression models using leave-one-out validation such that one model was trained for each participant using all of the other participants' data. We report the average error across all folds according to the MAE, limit of agreement, and Pearson's correlation coefficient (r). We further contextualize the results by reporting the accuracy with which FeverPhone is able to discriminate febrile participants. Rather than directly training a classification

Model Type	MAE	Pearson's r	Limit of Agreement	Sensitivity	Specificity
Dummy regression	0.452 °C	-0.884	1.017 °C	0.000	1.000
Linear regression	0.320 °C	0.621	0.754 °C	0.750	0.857
Quadratic regression	0.367 °C	0.557	0.897 °C	0.625	0.857
Random forest	**0.229 °C**	**0.763**	**0.731 °C**	**0.813**	**0.904**

Table 3.2: A table of regression and classification metrics for various model architectures. All of the models were trained using all four features based on the first 90 seconds of sensor data.

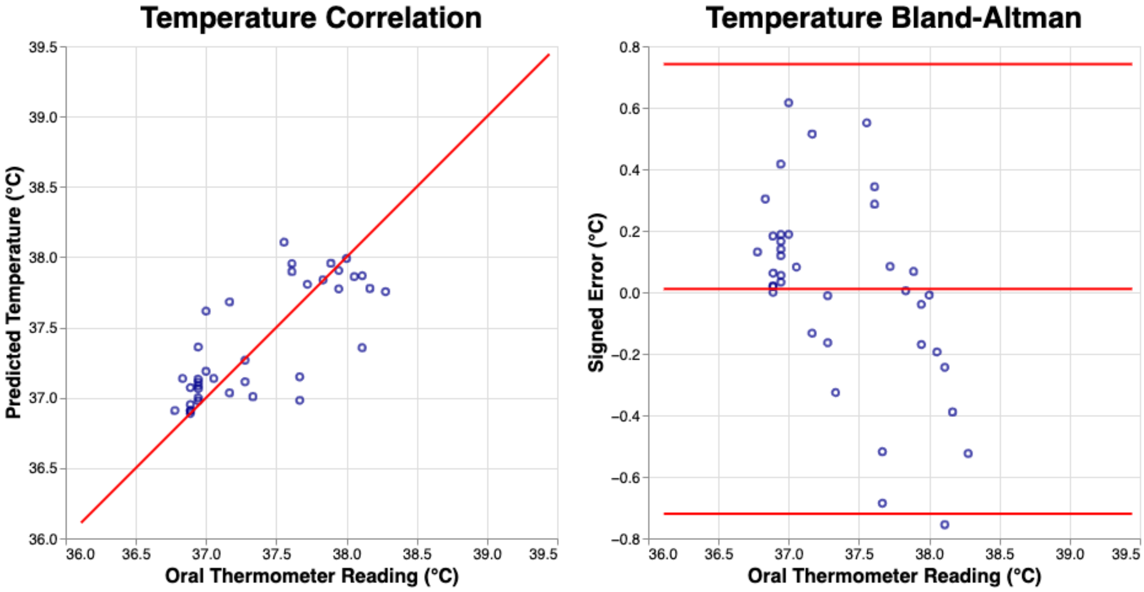

Figure 3.12: The (left) correlation and (right) Bland-Altman plot for the temperature estimates made by the random forest regression model. The models were trained using all four features based on sensor data that was collected over 90 seconds. The red lines on the Bland-Altman plot indicate the 95% limit of agreement.

model that could have different behavior from the regression model, we instead report classification accuracy by binarizing the ground-truth temperatures and FeverPhone's estimates according to the threshold for low-grade fever (≥37.5 °C) and then comparing their labels. To demonstrate the utility of the various FeverPhone model configurations, we also include the results of a dummy regression model that simply guesses the median of the dataset.

3.2.12 Results

RQ1.1: Model Selection

We tested three traditional regression models suited to handle low-dimensional data with varying degrees of nonlinearity: linear regression, quadratic regression, and random forest regression. The models were trained using all four features based on sensor data that was collected over 90 seconds. The results of this analysis are presented in Table 3.2.

Random forest regression outperformed the other options with respect to both regression and classification. Although there were linear relationships between the variables explored during our lab simulation study, the idealized conditions did not replicate the nonlinear nature of heat transfer that is inherent to the human body. This is one potential explanation for why random forest regression outperformed the other models. The ability of the random forest regression to estimate core body temperature is further illustrated in Fig. 3.12. The model achieved an MAE of 0.229 °C, a Pearson's r of 0.763 ($p < .001$), and a limit of agreement of 0.731 °C. Framing the performance of the random forest regression in terms of fever classification, the model achieved a sensitivity of 0.813 and a specificity of 0.904. These results are comparable and occasionally superior to some consumer-grade peripheral thermometers for both temperature regression and fever detection according to findings by [58] and [59].

RQ1.2: Duration of Data Collection

To investigate the optimal duration of contact between the user and the smartphone, we simulated different durations by restricting feature extraction to varying amounts of sensor data. The models were trained using random forest regression and all four features. The results of this analysis are presented in Table 3.3.

Increasing the duration of data collection generally improved the accuracy of FeverPhone and decreased the limit of agreement, which follows the intuition that more sensor data should lead to a more accurate representation of the thermistor's temperature curve. However, the performance started to degrade when the duration of data collection approached 120 seconds. This is likely due to the fact that the linear approximation we used for Newton's Law of Cooling was no longer applicable at these longer durations. We found that restricting feature extraction to the first 90 seconds of contact resulted in optimal results. This duration is shorter than what is required for mercury-in-glass thermometers and non-predictive digital oral thermometers, which can take between 30 seconds and 10 minutes to produce reliable estimates [121]. Although infrared temporal thermometers and consumer-grade digital oral thermometers with built-in prediction algorithms can make estimates faster, it is important to

Duration	**MAE**	**Pearson's r**	**Limit of Agreement**	**Sensitivity**	**Specificity**
Dummy regression	0.452 °C	-0.884	1.017 °C	0.000	1.000
30 sec	0.417 °C	0.224	0.951 °C	0.375	0.714
60 sec	0.324 °C	0.533	0.822 °C	0.625	0.809
70 sec	0.250 °C	0.717	0.764 °C	**0.813**	0.857
80 sec	0.236 °C	0.747	0.724 °C	**0.813**	0.857
90 sec	**0.229 °C**	**0.763**	0.731 °C	**0.813**	**0.904**
120 sec	0.240 °C	0.743	**0.691 °C**	**0.813**	**0.904**

Table 3.3: A table of regression and classification metrics for models trained with features extracted from varying amounts of sensor data. All of the models were trained using random forest regression and all four features.

note that these devices can still suffer from high error [58].

RQ1.3: Impact of Features

We explore the importance of various subsets of features. We first examine the performance of FeverPhone based on the primary feature — the rate of change in the thermistor's temperature — on its own. We then examine how adding supporting features from the thermistor or the touchscreen impact the model's performance. The models were trained using random forest regression based on sensor data that was collected over 90 seconds. The results of this analysis are shown in Table 3.4.

The model with all of the features led to the best performance as expected, but the baseline model with just the primary feature was not necessarily the worst performing configuration. Adding the supporting features from the touchscreen data improved the model, which was expected since the proximity to the thermistor and the amount of contact both control the amount of heat transferred from the user to the thermistor. However, adding the initial thermistor temperature as the only supporting feature degraded model performance with respect to metrics like MAE and AUC. This indicates that the touchscreen-based features provide important context to the thermistor readings and that there is an interaction between the supporting features that improves performance only if they are both included.

Although the best performing model in this experiment included all of the features, we anecdotally explored various combinations of features in other configurations. Most notably, we found that while the performance of the model including both thermistor and touchscreen-

Feature Subset	MAE	Pearson's r	Limit of Agreement	Sensitivity	Specificity
Dummy regression	0.452 °C	-0.884	1.017 °C	0.000	1.000
ROT + IT + CA + CL	**0.229 °C**	**0.763**	**0.731 °C**	**0.813**	0.904
ROT + CA + CL	0.251 °C	0.736	0.746 °C	**0.813**	0.904
ROT + IT	0.311 °C	0.610	1.028 °C	**0.813**	**1.000**
ROT	0.309 °C	0.582	0.979 °C	**0.813**	0.904

ROT = rate of change in the thermistor's temperature; IT = initial thermistor temperature; CA = contact area; CL = contact location.

Table 3.4: A table of regression and classification metrics for different feature sets. All of the models were trained using random forest regression with features based on the first 90 seconds of sensor data.

based features did not improve with interaction durations approaching 120 seconds, the model trained on just thermistor-based features did. This result indicates that the addition of touchscreen-based features allowed the model to be robust over shorter durations, whereas in the absence of these features, the model benefited from longer interactions.

RQ1.4: Fever Detection without Labeled Samples

In Section 3.2.7, we found that FeverPhone requires significant calibration to work on new devices. Although it is best to recollect a new dataset for each phone model, it can be difficult to acquire new training samples from febrile patients. Anomaly detection has been used in past work to distinguish healthy data samples from unhealthy ones in domains ranging from healthcare to mechanical engineering [122, 123], so we explored the potential of using anomaly detection for unsupervised fever classification. For FeverPhone, we assume that data samples from afebrile patients are more readily available during the training of a new phone model. Rather than fitting all four of FeverPhone's features to core body temperature using this dataset, we suggest the training of a model that maps the 3 supporting features to the rate of heat transfer — the primary feature that is correlated with core body temperature. This procedure generates a regression model that estimates the rate of heat transfer for healthy individuals. When this model is used for testing, the predicted rate of heat transfer is compared to the rate that is actually measured by the thermistor. If the difference between the two exceeds a predefined threshold, then we surmise that the rate of heat transfer has been significantly mischaracterized and that the user likely has a fever provided that there is sufficient coverage in the training dataset.

We evaluated this approach using the FeverPhone configuration selected for our previous research questions, only this time by training a model to estimate the rate of heat transfer from the three supporting features. This model was trained using 5-fold cross-validation on all the data from patients without a low-grade fever. Using a threshold of -0.001 to classify samples as anomalies and testing this model on the febrile patients, we found that this unsupervised approach achieved a sensitivity of 1.00 and specificity of 0.713.

Summary

To summarize our analyses, the optimal configuration of FeverPhone uses a random forest regression model trained on features from the thermistor and the capacitive touchscreen to describe the rate of heat transfer from the forehead to the smartphone. These features are optimally calculated over a 90-second period, which provides a suitable amount of time to accurately characterize the rate of heat transfer while still obeying assumptions related to linearity. In this configuration, FeverPhone achieved an MAE of 0.229 °C, a limit of agreement of ±0.731 °C, and a Pearson's r of 0.763. When those readings are used for fever classification, FeverPhone achieved a sensitivity and specificity of 0.813 and 0.904 respectively. We also found that it is possible to detect low-grade fevers without training a model on such instances. Using unsupervised anomaly detection, we were able to identify febrile patients with sensitivity and specificity of 1.00 and 0.713 respectively.

3.3 Discussion

This work serves as a foundation for core body temperature measurement using unmodified smartphones. Our goal is not to replace existing thermometers but rather to develop an accessible alternative for cases when traditional thermometers are not readily available, namely telehealth consultations and resource-constrained environments. Nevertheless, we demonstrated that FeverPhone achieves accuracy levels that are comparable to some consumer-grade peripheral, tympanic, and axillary thermometers. In the discussion that follows, we elaborate upon the impact of our findings, considerations for future work, and the limitations of our research.

3.3.1 Key Findings

The design of FeverPhone is grounded in Newton's Law of Cooling, the same working principle that underlies most clinical-grade thermometers. However, using this equation alone was not enough to make FeverPhone feasible, which led to two innovations. First, we found that properly characterizing the exponential rate of heat transfer requires as long as 8 minutes of data collection, which would be unsuitable for human interaction. Our results demonstrate that approximating Newton's Law of Cooling using a linear function over a shorter period can lead to satisfactory results. Second, we observed that the region of contact between the user and the smartphone's screen can vary according to how they position the phone against themselves and the curvature of the forehead itself. Since the contact surface influences the rate of heat transfer from the user to the thermistor, our feature ablation study shows that incorporating features representative of it improves the accuracy of FeverPhone. These two innovations should be useful to any future systems that would rely on direct contact between the phone and another object for temperature estimation.

3.3.2 Supporting Practical Scalability of FeverPhone

One of FeverPhone's supporting features relates to the weighted centroid of the region of contact between the user and the smartphone as measured by the touchscreen. Conceptually, this feature was designed to encode information related to the distance between that region and the thermistor within the phone. When the thermistor's position is fixed and known, the distance can be calculated using a uniform offset. However, the same offset cannot be applied to other phones since their thermistors are likely in different locations, making this feature in particular difficult to generalize across smartphone models. The distance could be directly calculated if smartphone manufacturers were to expose the location of their thermistors to model developers, which we believe would dramatically improve the generalizability of FeverPhone. It would also be beneficial for FeverPhone if smartphone operating systems were to provide an API for throttling temperature-sensitive tasks like network utilization and momentarily increasing the sensor resolution. Although we found the optimal duration of data collection to be 90 seconds, we believe that a shorter duration could be feasible given a more precise and higher resolution thermistor. Smartphone operating systems could also provide an API for directly reading raw touchscreen capacitance and thermistor values to make FeverPhone and similar applications more ubiquitous. Finally, manufacturers could provide open-access data on characteristic thermal curves for the thermistors on their devices, which would allow for faster model generation. This data could be generated by logging thermistor data during factory stress tests. With enough data across phone models and thermistors, a more general model could be created based on thermistor location as well as the phone's material, size, and mass. Having such a model would significantly improve the ability to generalize FeverPhone across smartphone models.

3.3.3 Estimating Core Body Temperature from Multiple Locations

The mapping between external and core body temperature could be facilitated by applying the working principle of FeverPhone to other smart devices such as smartwatches or other body-mounted wearables. Although we selected the forehead for ergonomic and stability reasons, future investigations may find other testing sites to have their own advantages. Leveraging multiple testing sites across the body could also yield performance benefits. Beyond providing redundant measurements that can be compared and averaged, thermoregulation may vary across body sites in a way that reveals new information about the user's core body temperature. Combining measurements across the body could also serve as an indicator of ambient air temperature or user activity as the surface temperature of different body sites have been shown to be impacted by those factors at different rates [112, 124]. Despite these opportunities, idiosyncrasies in people's physiology may require user-specific modeling to properly reconcile measurements across the body, so we leave this opportunity to future work.

3.3.4 Limitations

Although we explored a number of confounds in both our lab validation and clinical studies, there were other confounds in that latter that were not characterized. For example, participants used FeverPhone in varied postures due to the nature of their medical ailments. Some participants were required to be upright while others had to be reclined as part of their treatment, which may have impacted their thermoregulation during our trials. Trials were also conducted in varying locations of the emergency department (waiting room, hallway, patient room). We explored the impact of ambient air temperature in our lab validation study, but subtle changes across the emergency department may have contributed to the variance in our clinical study.

Another limitation to our study was the range of core body temperatures included in our study. We focused specifically on FeverPhone's performance on people who had nominal (36.6 °C–37.2 °C), abnormal (37.2 °C–37.5 °C), low-grade fever (37.5 °C–38 °C), or borderline fever (∼38 °C). People with core body temperature within these ranges are the most likely to be unsure of their condition and could benefit from a fever screening technology. We omitted participants with extremely severe fevers (>38.6 °C) since the need for fever screening is less important for these individuals and commercial thermometers are known to have reduced accuracy when people have sweaty or clammy skin [118, 119, 120]. Nevertheless, future work may want to consider additional features to identify these potential confounds so that FeverPhone can better support people with severe fevers.

Lastly, FeverPhone uses root-level access for reading a broad set of thermistors and a custom user-debug Android kernel for accessing raw touchscreen capacitance data. These requirements limit FeverPhone to phones with fine-grained sensor access. Temperature data from the battery thermistor is more readily accessible but did not yield comparable performance relative to temperature data from other thermistors. Still, future work could explore a more accessible version of FeverPhone with specific considerations paid to the battery thermistor. Similarly, the `TouchEvent` API is publicly accessible at the cost of coarse-grained information about touch location. The use of this API in place of the raw capacitance data could be explored in future work.

3.4 Conclusion

In this chapter, we presented FeverPhone, a system for estimating core body temperature and identifying low-grade fevers using a commodity off-the-shelf smartphone.

We address **RQ.1**: What is a signal which can be collected from sensors onboard commodity smartphones with no additional hardware for detection of ILI infection at scale and how can it be utilized most effectively? By developing a fever monitoring system leveraging just the sensors integrated into the device. We also demonstrate how a simple interaction – by bringing the smartphone in contact with the forehead – can re-contextualize this sensor to most effectively collect core body temperature by measuring heat transfer from the skin at a thermal site which is both easily accessible and highly correlated with core body temperature.

FeverPhone adheres to the same constraints as existing contact-based thermometers and has comparable accuracy to some consumer-grade temporal, tympanic, and other peripheral thermometers. However, FeverPhone only relies on the smartphone's built-in thermistors and touch screen to serve this functionality. In a clinical trial with human participants of nominal and elevated core body temperature, we found FeverPhone to have an MAE of 0.229 °C, a limit of agreement of ±0.731 °C, and a Pearson's correlation coefficient of 0.763. When those readings are used for fever classification, FeverPhone achieved a sensitivity and specificity of 0.813 and 0.904 respectively. We believe that, if deployed at scale, FeverPhone has the potential to improve remote care and even epidemiology by providing an accessible method of temperature screening to smartphone users.

Chapter 4

Wearable-based Biomarkers of Immune response for Presymptomatic Detection of Influenza

In this chapter, we discuss the use of commodity wearable smartwatches for assessing physiological biomarkers of H10N7 "Bird Flu" influenza in its earliest stages, with the goal of identifying influenza cases before traditional symptoms arise. We do this through a two week observational study of 34 patients challenged with H10N7 monitored by commodity wearable devices (Google Fitbit Sense 2). During this study, each participant is inoculated through nasal inspiration with the influenza virus and continuously monitored using a rotating set of smartwatch devices until they fully recover. Results show that measured resting heart rate, skin temperature, and electrodermal activity increase across infected individuals, while heart rate variability decreases and that these effects are more pronounced for the participants who experienced some symptoms and produced a positive flu test at least once during the study duration. However, these biomarker trends were observed even for participants without positive flu tests and as early as the very first day of infection, indicating that commodity wearable devices may provide an indication of viral infection for traditionally undetectable cases or prior traditionally measured symptoms.

4.1 Introduction

Estimates from the CDC suggest that 35 million people contract influenza each year and more than half make at least one visit to a healthcare provider [125]. These preventable communicable diseases can spread quickly through peer-to-peer interaction, consuming healthcare resources which has cascading effects on the quality of care throughout the system [11]. Stopping the transmission of infectious disease is therefore most immediately achieved through limiting interaction between infected and uninfected individuals. However, this depends on an awareness that one is infectious, and many viral infections become transmissible before the onset of human perceptible symptoms [126]. Additionally, traditional means of validating infectious disease cases such as PCR tests are further lagging indicators, typically peak in effectiveness an additional 3 days after symptom onset [127]. This delay in both symptom onset and diagnosis leaves a critical window in which individuals may be contagious yet asymptomatic or unaware of their infection, leading to opportunities for further spread. Conversely, delay in diagnosis due to reduced symptom presentation is associated with increased mortality in influenza cases [128]. Therefore, reducing both the spread and the severity of influenza and other infectious disease depends on early detection of viral infection during this critical pre-symptomatic window using leading indicators that precede traditional symptoms. This early detection is particularly important for viruses which have long incubation periods where infected individuals may be asymptomatic yet still contagious or infections which may present entirely asymptomatically.

The widespread adoption of wearable health monitors is a promising approach for disease prevention with a growing 1-in-4 of Americans using consumer wearable fitness trackers [2]. These devices provide continuous health monitoring by tracking cardiovascular biosignals such as resting heart rate, heart rate variability (HRV), and other signals relevant to infectious disease such as skin temperature and electrodermal activity (EDA). Prior investigation by Natarajan et al. [25] into these wearable biosignals has shown that measurable changes in cardiovascular signals such as increased resting heart rate and decreased HRV present multiple days before the onset of symptoms for smartwatch users who tested positive for Covid-19. However, naturalistic retrospective studies like this one lack knowledge of the time of infection, making it impossible to determine how soon after exposure these biomarkers respond.

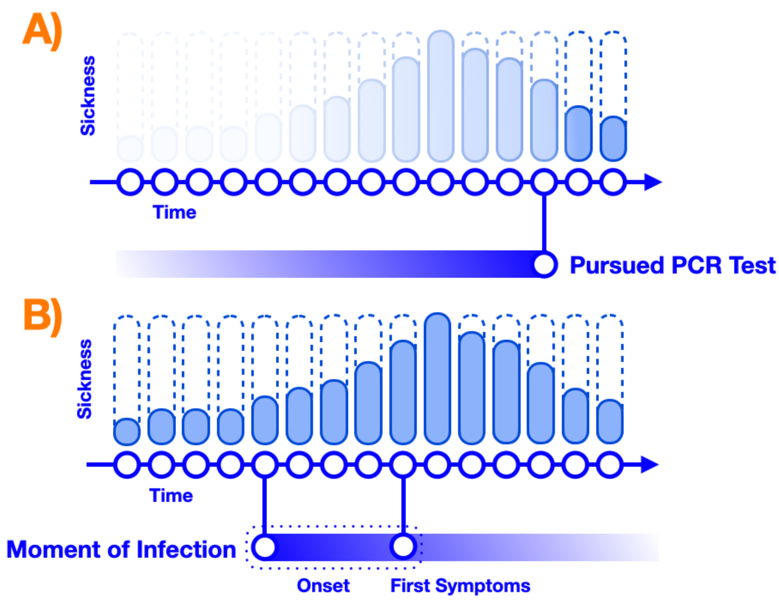

Figure 4.1: A) An abstract illustration of flu severity from the perspective of a retrospective analysis where illness becomes increasingly plausible until a PCR test is taken to confirm. B) An abstract illustration of flu severity from the perspective of a prospective analysis in which the moment of infection is known and severity can be cleanly mapped to the days post-infection and pre-recovery.

In the field of immunology, disease progression is assessed through prospective challenge study in which participants are "challenged" with a particular virus through intentional exposure at some known time and observed as the disease progresses [129]. This approach provides ground truth knowledge of how the disease impacts the observation metrics across time relative to the known moment of infection. The difference between retrospective and prospective analyses is illustrated in Figure 4.1.

In order to elucidate the effects of ILI on measurable wearable biomarkers, we conduct a 34 participant influenza challenge study using the novel H10N7 "Bird Flu" virus, a mild influenza virus with typically asymptomatic presentation. During the study the participant's health is monitored continuously via commodity wearable fitness trackers (FitBit Sense 2) as well as discreetly with traditional vital sign assessments, routine PCR tests, daily symptom surveys, and bloodwork. We show that over the course of the study the majority of participants exhibit elevated resting heart rate, skin temperature, and skin conductance and reduced HRV both as early as the first night after exposure and that this is true even for both fully asymptomatic cases and participants who did not produce a single positive PCR test. This indicates that wearable smartwatches have the potential to detect infectious disease both early and for very mild cases.

4.2 Study Overview

To conduct this study we recruited 34 participants across 7 cohorts of at most 5 simultaneous participants from October 2023 to November 2024. The study was conducted at the National Institute of Allergy and Infectious Diseases (NIAID) facility – part of the National Institutes of Health (NIH) Clinical Center in Bethesda Maryland. Each participant remained at the facility in isolation for either 9 days or until they produced two negative PCR test results if they remained PCR-positive on the final day. Participants were initially consented and screened for pre-existing infection through PCR tests upon arrival at the clinical center on day -1 (referred to as the baseline day) and were inoculated with H10N7 influenza on the morning of day 0. The participants were then tested for any existing viral infections and if a participant's PCR results returned positive on day -1 for any of the tested viral infections, they were excluded from the current cohort but were invited for following cohorts. This ensures biomarkers on the baseline day reflect nominal health conditions, increasing the signal of relative changes in biomarkers on or after day 0 due to immune response to H10N7.

Participants were challenged with increasingly large doses of the virus starting at 10^4 TCID50 in the first cohort increasing to 10^7 TCID50 in the final 4 cohorts. The dose increased by an order of magnitude each cohort until the majority of participants in a single cohort produced a PCR-positive. Only participants with 10^6 and 10^7 TCID50 doses produced PCR-positive results.

Traditional Metrics: On all days of the study, including the baseline day, participants completed a Patient-Reported Outcome (FLU-PRO) survey, a standardized and validated questionnaire for evaluating influenza severity [130]. Participants also provided a daily nasal swab sample for PCR tests everyday excluding day 0, the day of inoculation, due to residual virus from intranasal inoculation. Additionally, participants received 4 vital assessments a day every 4 hours including measuring temperature, respiration rate, blood pressure, mean arterial pressure (MAP), blood oxygen saturation (SpO2), and heart rate measured by traditional pulse check and EKG. Blood work was used to measure genetic expression as a measure of immune system response and was captured four times per participant on days -1, 2, 4, and 6.

Continuous Wearable Biomarkers: Participants were also provided with a set of three Fitbit Sense 2 smartwatches which they rotated approximately every 10 hours to match the device's battery capacity (once at 10pm and again at 8am) to ensure continuous data collection throughout the study. These devices measured: heart rate via PPG, physical motion from an accelerometer, skin temperature, and EDA (i.e., skin conductance). These signals were recorded at 25 Hz, 50 Hz, 1 Hz, and 20 Hz respectively. The data was manually pulled from the smartwatch once per day by a clinical on-site collaborator due to limited on-device memory. The 3 identical devices were allowed to rotate in a round-robin fashion between 1) being worn by the participant until the battery was depleted, 2) in the hands of the collaborator for off-loading data, and 3) charging in the patients room to facilitate a timely swap once the current watch battery was drained.

4.2.1 Extracting Heart Metrics from PPG

While the skin temperature, EDA, and motion sensor all measure their desired output signal directly, the PPG sensor requires some processing to extract meaningful cardiovascular metrics. We will briefly cover our approach. In order to extract heart rate and HRV metrics from the raw PPG timeseries, we first apply a band pass filter from 0.6 Hz to 3 Hz to limit the range to the expected resting heart rate. We then apply the SciPy $find\_peaks$ algorithm to identify candidate systolic peaks in the PPG timeseries. We then calculate the time series of differences in time delay between successive systolic peaks – known as the RR-interval or Normal-Normal – which directly reveals heart rate. In order to filter out noisy regions of PPG signal, we exclude regions where the ratio of successive RR-intervals are greater than 30% difference. i.e., we removed peak RR_0 where the following condition was met $(RR_0 - RR_1) - (RR_1 - RR_2) > 0.3 * (RR_0 - RR_1)$. This leverages the expectation that resting heart rate variation is smooth, unlike heart rate while awake which may experience transient changes due to emotional response or activity. This addresses cases in which both the systolic and diastolic peak are detected in peak finding due to a steep dicrotic notch. This is important as the time between systolic and diastolic peaks is often much shorter than the time between successive systolic peaks and ensures the accuracy of downstream HRV metrics which are calculated on this peak-to-peak distance. To capture resting heart rate metrics, we only analyze data from 0:00 to 5:00 while participants are sleeping. Measurements were then aggregated using a sliding window at the 5-minute scale and discarded if within-window coverage was below 70% (i.e., less than 70% of the 5 minutes was valid data which did not meet our 30% RR successive difference threshold). Average resting heart rate was then calculated as the mean of the RR-intervals / 60 seconds and HRV was calculated as the standard deviation of normal-normal (SDNN) (normal-normal interval is equivalent to RR-interval) across each window. All other metrics from the watch (i.e., skin temperature, EDA, and motion were directly revealed by the sensor signal.

4.2.2 Processing Blood Work and Genetic Expression

The blood work is processed by a wet lab to return gene expression across approximately 30,000 gene symbols. The gene expression on each day for each participant is then processed using Single-Sample Gene Set Enrichment Analysis (ssGSEA) [131] via `gseapy` against a set of common gene pathways to generate pathway level enrichment scores. These scores capture the biological processes triggered in the body of the participant on each day that blood was drawn. There are many possible genetic pathways to test, but specific pathways such as the Hallmark IFN (i.e., $HALLMARK\_INTERFERON\_ALPHA\_RESPONSE$ and $HALLMARK\_INTERFERON\_GAMMA\_RESPONSE$) [132] represent interferon signaling and are commonly used for measuring the antiviral arm of immune response in the blood [133] and can serve as additional grounding for our mild, possibly asymptomatic, flu cases where PCR detection thresholds may not be met or corroborate the cases where participants had sufficient viral load to produce a positive flu test.

4.3 Results

Now we will discuss the results of the challenge study, starting with the clinical test results and following with the digital biomarker results for different meaningful groupings of participants.

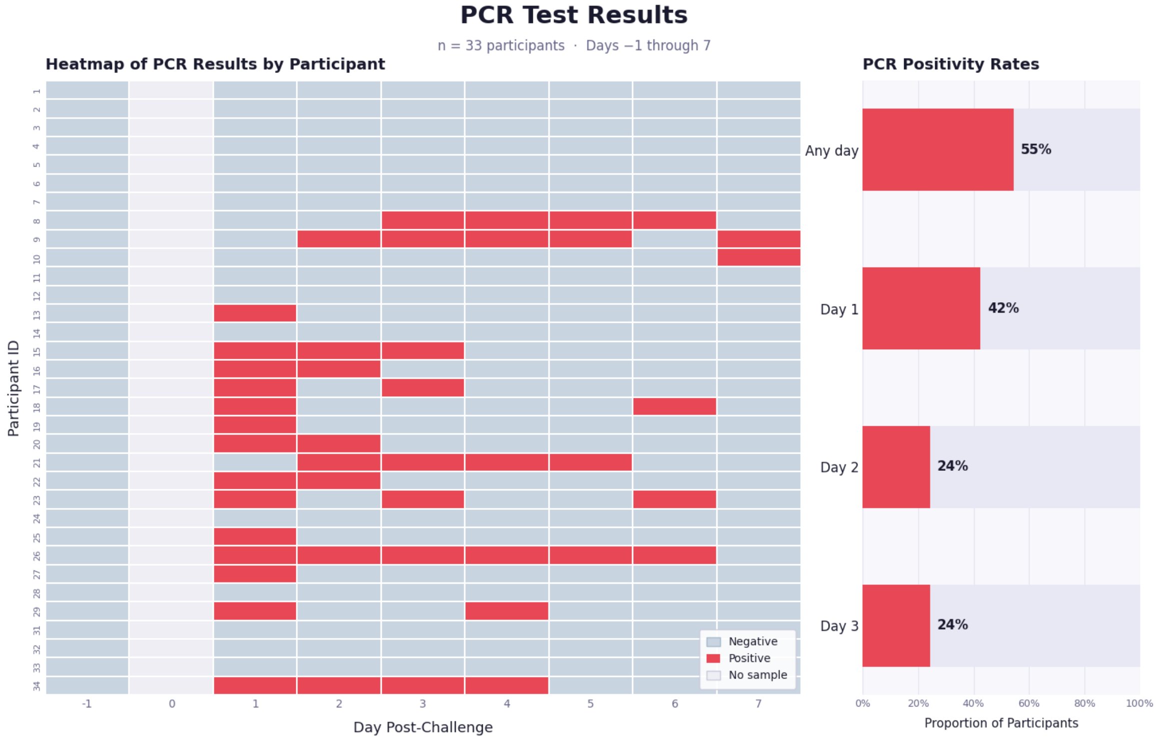

Figure 4.2: A table showing the study days for which participants received positive PCR test results from day -1 to day 7. No PCR test was administered on day 0 due to a confounding factor of nasal inspiration of the virus on that day. PCR positive rates are shown to the right across the full study and across the most significant early study days.

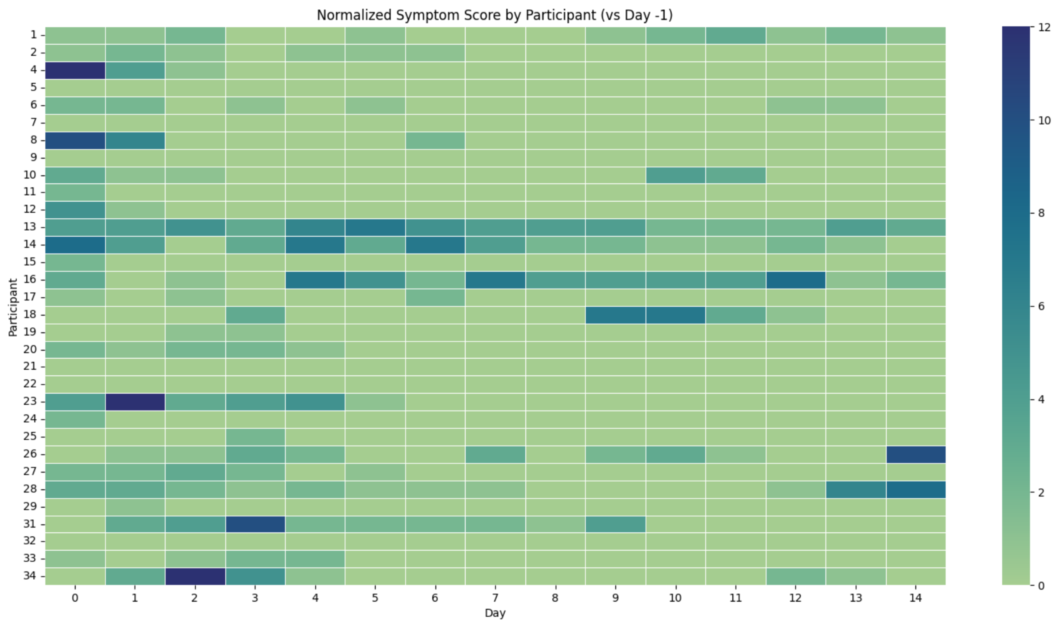

Figure 4.3: A table showing the weighted sum of self-reported symptom severity by participant on each day of the study where each symptom contributes a score of 1-5 based on the self-reported severity on a likert scale.

4.3.1 Clinical Test Results

PCR, Symptoms, and MMID: Figure 4.2 shows the PCR results for each participant on each day of the study and Figure 4.3 shows the self-reported symptom count for each participant for the 2 weeks following inoculation with H10N7. During the challenge study 18/34 (53%) participants received at least 1 positive PCR test, while 26/34 (76%) participants reported onset of flu-like symptoms post-infection beyond the symptoms they experienced on the baseline day. 6/26 of these participants reported experiencing some symptoms before inoculation. None of our participants experienced a fever ($>38°C$) or even low-grade elevated core body temperature ($>37.5°C$). 15/34 (44%) participants reported both onset of flu-like symptoms and received a positive PCR test during the study, meeting the clinical definition for mild-to-moderate influenza disease (MMID). This means 11/34 (32%) participants were symptomatic but never produced a viral load great enough to receive a positive PCR test and 3/34 (9%) participants were PCR positive while fully asymptomatic and 4/34 (12%) participants were neither symptomatic nor produced a PCR positive. Of the 15 participants with MMID, all but 2 were within the cohorts the received the maximum viral dose of H10N7 and those remaining 2 were in the second-to-largest dose cohort. Of the 21 participants in the max dose cohorts, 15/21 (71%) produced a PCR positive and 13/21 (62%) were MMID.

All participants were infected with H10N7 and experienced some kind of immune response, however those participants in the later cohorts with a higher dose expectedly experienced a stronger immune response leading to more symptoms and measurable signals in genetic expression. We therefore report resulting biomarkers across subgroups of participants such as "shedders" or those who produced at least one positive PCR test or MMID participants to reflect those who are most likely to exhibit a measurable change in wearable biomarkers. While these groupings capture the biomarkers characteristic of more clear flu infection, we also capture the biomarkers of non-shedders and non-MMID participants to reflect the more asymptomatic cases.

Genetic Expression: After applying ssGSEA to each blood sample, the enrichment scores for interferon pathways are aggregated across viral shedders and non-shedders (i.e., those with and without at least one positive PCR test during the study respectively) to show increased interferon pathway enrichment within the shedding group. The resulting enrichment scores are shown in Figure 4.4. The average interferon pathway enrichment score ranks near the top of the list of sorted enrichment scores across the Hallmark set indicating that the gene expression of these participants aligns with aspects of immune response to viral respiratory infection. This serves as additional grounding that the H10N7 virus produces immune system activity, even for some participants who do not produce a positive PCR test or experienced symptoms.

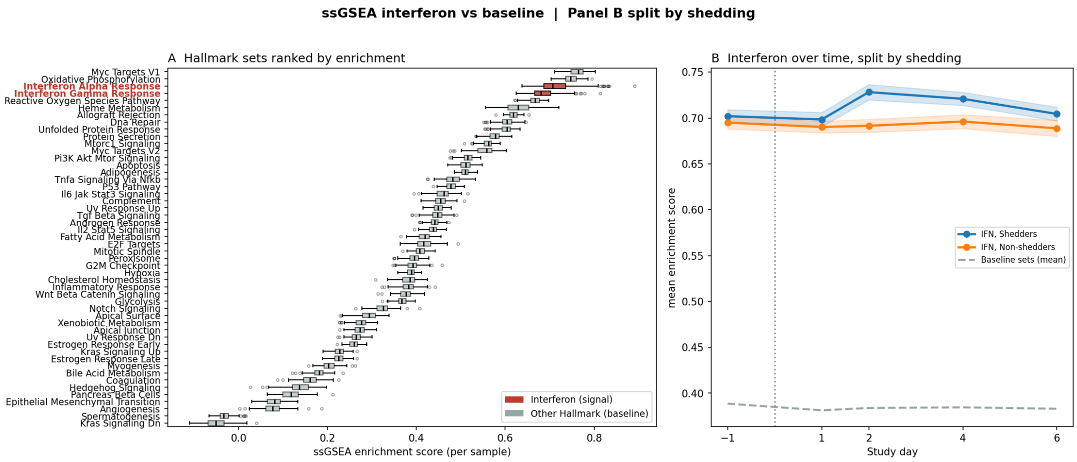

Figure 4.4: A) The Hallmark pathways sorted by enrichment across all study days and participants. B) the interferon pathway enrichment scores aggregated across shedders and non-shedders on the 5 days blood work was taken

Table of Participant Indicator Variables: The table of all indicator variables used to group participants by immune response or illness severity is shown in Table 4.1.

4.3.2 Digital Biosignal Trends

The average relative change in the four biosignals collected from the wearables during the post-infection study days reflect biologically plausible effects of immune response. The high

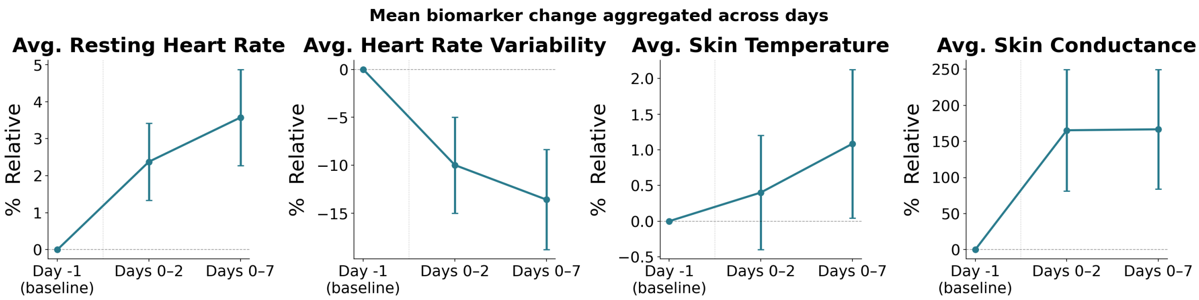

Figure 4.5: The relative change compared to baseline in all four digital wearable biosignals aggregated across all infected participants and aggregated across time (both the first 3 days of the study and across all days post-infection). Error bars show standard measurement error.

Table 4.1: Per-participant infection indicators. Symptoms = ≥ 1 symptom(s) risen above the participant's own day -1 baseline (N_{sym} = peak such count); ΔIFN = interferon ssGSEA score at day 2 minus day -1; High IFN = ΔIFN > non-shedder mean + 1 SD (threshold = 0.01).

Participant	Shedder	Symptoms	N_{sym}	MMID	$\Delta\text{IFN}_{d2-d(-1)}$	High IFN
1		✓	2		N/A	N/A
2		✓	2		N/A	N/A
3			0		-0.03	
4		✓	17		N/A	N/A
5			0		-0.02	
6		✓	2		+0.01	
7			0		-0.02	
8	✓	✓	8	✓	-0.00	
9	✓		0		+0.15	✓
10	✓	✓	3	✓	+0.01	
11		✓	1		+0.01	
12		✓	5		+0.01	
13	✓	✓	7	✓	+0.01	✓
14		✓	8		+0.01	
15	✓	✓	2	✓	+0.10	✓
16	✓	✓	4	✓	+0.02	✓
17	✓	✓	2	✓	-0.01	
18	✓	✓	3	✓	-0.00	
19	✓	✓	1	✓	-0.01	
20	✓	✓	2	✓	+0.01	✓
21	✓		0		+0.03	✓
22	✓		0	✓	-0.01	
23	✓	✓	10	✓	+0.08	✓
24		✓	2		-0.01	
25	✓	✓	1	✓	+0.01	
26	✓	✓	4	✓	+0.02	✓
27	✓	✓	3	✓	+0.01	
28		✓	7		-0.01	
29	✓	✓	1	✓	+0.02	✓
31		✓	8		+0.02	✓
32			0		+0.00	
33		✓	2		-0.00	
34	✓	✓	6	✓	+0.06	✓

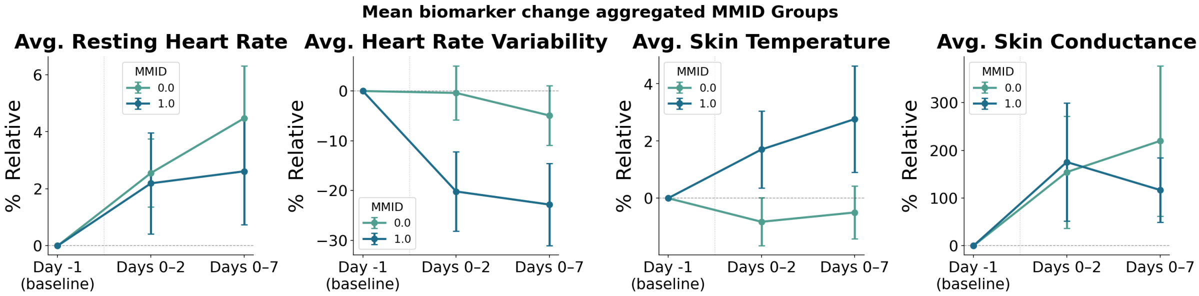

Figure 4.6: The relative change compared to baseline in all four digital wearable biomarker signals aggregated across MMID and non-MMID participants and aggregated across time (both the first 3 days of the study and across all days post-infection). Error bars show standard measurement error.

level trends in biosignals with respect to all participants in the study are shown in Figure 4.5, while participants grouped by illness severity via MMID status are shown in Figure 4.6. Trends per biosignal are discussed in detail below. While multiple labels such as symptomatic, viral shedding (i.e., PCR test results), MMID, and high immune response from gene expression analysis were tested for stratifying participants, given the significant overlap between these groups, only the results for MMID are shown. However, grouping participants by these other labels provided similar trends.

Resting Heart Rate: The average resting heart rate of our participants (both MMID and non-MMID) increased by 2% during the first 3 days of infection (including the day of inoculation) and persisted for the full duration of the study. This may indicate an increase in cardiovascular activity as a result of the immune system responding to the introduction of the virus to the body. The increase in resting heart rate for the non-MMID participants indicates that this effect may be measured for asymptomatic cases or those with viral loads below PCR detection thresholds.

Heart Rate Variability: The average HRV of only the MMID participants decreased, but decreased significantly by over 20% in just the first 3 days post-infection. This may indicate that a reduction in HRV measured during sleep could provide an early signal of immune response to more severe infections that triggers symptoms and passes the PCR detection thresholds.

Skin Temperature: The average skin temperature of only the MMID participants increased, with an increase of 2% over baseline. This may indicate that an increase in skin temperature

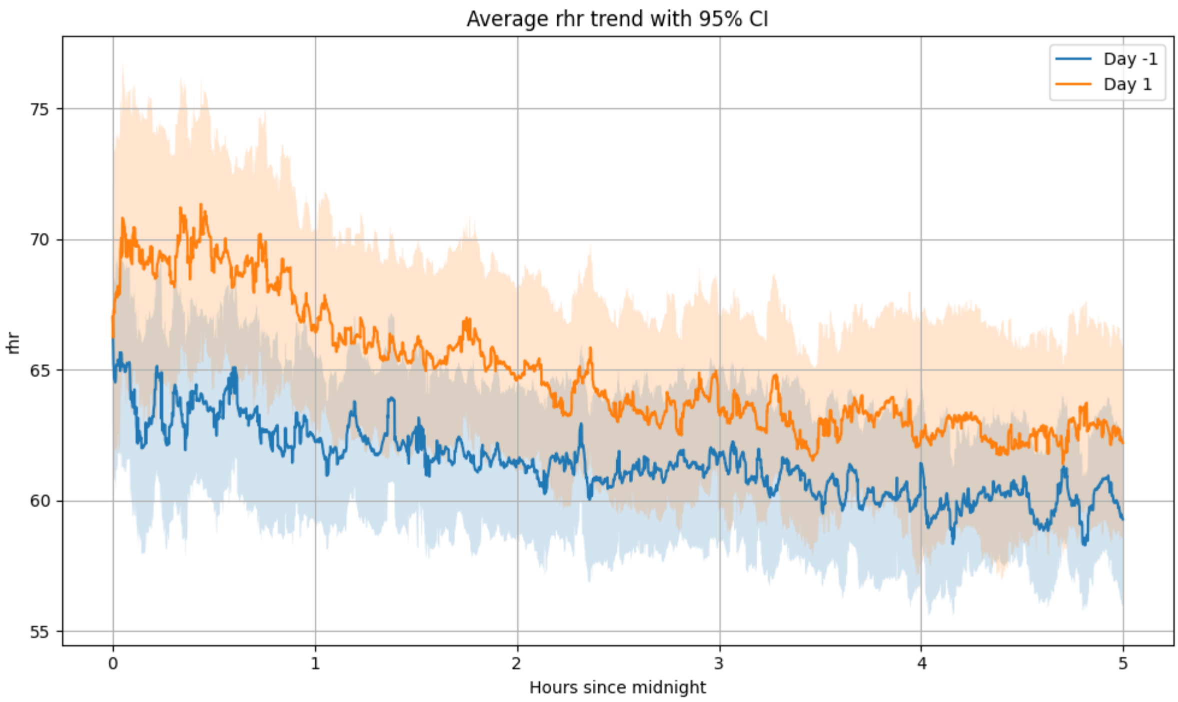

Figure 4.7: Average participants resting heart rate on day -1 before infection and on day 1 after infection at the 5-minute scale between midnight and 5am. The average participant's difference in resting heart rate after becoming infected is greatest in the early morning between midnight at 3am.

could be used as an early warning sign of immune response to more severe infections that trigger symptoms and pass the PCR detection threshold.

Skin Conductance: The average skin conductance nearly doubled in the first 3 days after infection for both the MMID and non-MMID participants. This may indicate that an increase in skin conductance may indicate immune response to subtle flu infections that may be asymptomatic or not pass the PCR detection threshold.

Prior research observed that flu symptom onset typically occurs 3 days after infection [127]. These results therefore indicate that the relative changes in biosignals measured from commodity wearables may be associated with exposure to H10N7 in the earliest days of infection, even before symptom onset. Additionally, these changes observed in resting heart rate and skin conductance may be associated with exposure to H10N7 for subtle or even asymptomatic cases.

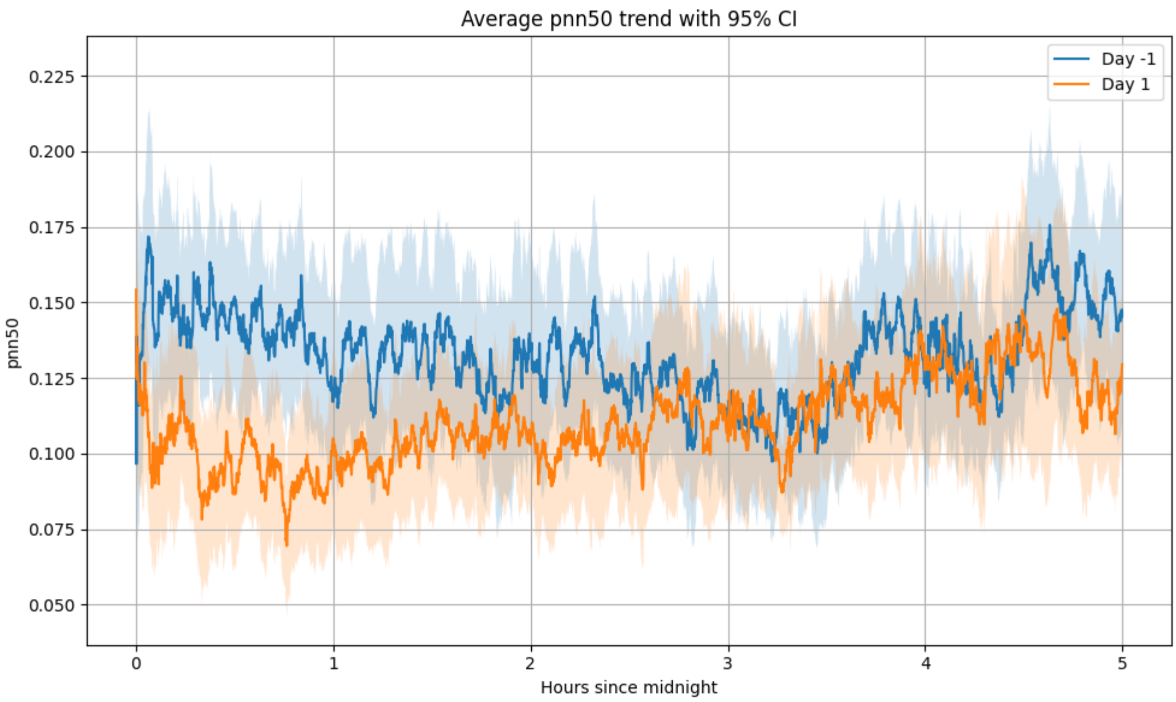

Figure 4.8: Average participants heart rate variability on day -1 before infection and on day 1 after infection at the 5-minute scale between midnight and 5am. The average participant's difference in heart rate variability after becoming infected is greatest in the early morning between midnight at 3am.

4.3.3 Within Night Comparison

Figures 4.7 and 4.8 show the resting heart rate and heart rate variability at a 5-minute scale between midnight and 5am averaged across all participants on the baseline day and the second night after infection (night after day 1). Figure 4.7 shows that the difference between resting heart rate during the night post-infection and the baseline night is most significant earlier in the night, although the difference is present across the full range. Figure 4.8 shows that the difference between heart rate variability during the night post-infection and the baseline night is most significant and only separable between the hours of midnight and 3am. Figures 23-26 show these same patterns on a 30-minute scale for all participants and the participants in the max dose cohorts respectively.

4.4 Discussion

We analyzed over 7,000 hours of continuous wearable health data for 34 participants across 8-10 days post-infection with H10N7 influenza as compared to 1 baseline day prior to infection and validated relative within-participant trends against self-reported symptoms and PCR tests. Due to the H10N7 influenza being a mild virus, the majority of participants showed little to no symptoms and many participants never produced a positive PCR test. The results of this study therefore demonstrate not just the presence of these digital biomarker signals for mild perceptible cases, but fully asymptomatic ones. Additionally, because these cases were mild and PCR tests were conducted daily, unconditional on the participants perceived symptoms, these results may additionally represent early biosignal trends exhibited prior to when a participant may have elected to pursue a PCR test on their own.

4.4.1 Advantages of Digital Biomarkers Compared PCR Tests

Despite all being exposed to H10N7 only approximately half of the participants received a positive PCR test during the duration of the study, with an average number of positive PCR days of 2.58 $\pm$ 1.97 days for those who received a positive (average PCR positive of 1.57 $\pm$ 1.99 for all participants). The majority of participants who produced a positive PCR test did so on the first day after infection (day 1) and 5 participants produced negative PCR tests on days between positive PCR tests. While PCR tests remain the gold standard of influenza surveillance, this highlights their variability as a spot estimate. PCR tests measure virus replicating in the nasal mucosa which may not directly reflect the infectiousness of the individual or their immune response directly. Additionally, they are measuring a single moment in time. Taken too early or too late, PCR tests may result in false negatives (i.e., taken before replication exceeds detectable thresholds) or a stale positive (i.e., taken right before replication drops as symptoms and infectiousness subside). Additionally, in the real-world, PCR tests are opt-in, meaning that there is no guarantee that the participants who received a positive PCR test on the first day of the study may have even pursued a PCR test in a natural environment as the mild symptoms they experienced may not have merited the time or cost investment necessary to get tested in-the-wild. Participants typically would opt for a PCR test if they had significant symptoms, meaning that only MMID participants are even likely to pursue a PCR test in-the-wild and given real-world opt-in testing, a subset of these participants may not have tested until well after symptom onset. This highlights the opportunities of passive wearable biosignals as complementary in ILI surveillance as they can capture data continuously and additionally measure something entirely different – the systemic immune response to virus in the body. Wearable ILI surveillance therefore suffers from entirely different behavior and systemic limitations which make them a complementary system to support PCR based surveillance.

Similarly, while PCR tests provide a strong indication that participants are sick during a severe case of influenza, it is possible for PCR tests to return false negatives for mild or asymptomatic cases. This is on top of the aforementioned survivorship bias where many mild or asymptomatic, yet contagious, cases of infectious disease may never result in the individual pursuing a PCR test at all due to lack of symptoms. Digital biomarkers from wearables not only offer a complementary signal to assess one's own health during infectious disease outbreaks or after exposure, but may also improve the utility of PCR tests for confirmation by introducing supplemental low-cost screening signals to pursue a test in mild or asymptomatic individuals. This could potentially increase the reliability of case counts in epidemiological models by addressing a subset of false negatives due to this behavioral bias in electing to flu test.

4.4.2 Limitations

A limitation of this study is that, unlike naturalistic study, participants are informed ahead of time that they will be exposed to an influenza virus. This can bias participants to be more critical of their responses to symptom surveys, confounding results. Additionally, participants are monitored only within the confines of the NIH facility. It is possible that this change of environment induces variation in some biomarkers such as heart rate or HRV due to stress or even their own perception of their symptom severity. This effect may be limited yet still present during sleep due to stress being a conscious effect which is less pronounced while participants are unconscious. Similarly, since PCR tests are routinely administered, it is impossible to assess whether the biosignal trends observed would have predated any opt-in PCR test results that users would take in-the-wild. This is a fundamental limitation of prospective control studies.

Another limitation of this study is the disparate clinical labels. While viral shedding, symptoms, and immune response each could be used as a ground truth signal on their own (and overlap in many cases), they each tell something distinctly different about the participant and suffer from unique biasing and errors. In this study, we defer to the NIH standard definition of MMID as a baseline ground truth to subset our patients into a "positive" class, but in theory the most directly correlated label to wearable biosignal trends should be the immune system response. However, because genetic pathway enrichment score must be binarized into positive and negative classes, there is a significant risk of overfitting this threshold to the study population, impacting generalizability of results, and with such a small study population, it is unreasonable to subset into a subset for deriving the enrichment score threshold from a holdout evaluation set.

4.4.3 Future Work

Future work can leverage the dataset sourced by this study for evaluating flu prediction models for early detection. Specifically, the data from this study allows for evaluation of models in terms of time-since-exposure as a new evaluation metric for ILI detection along a timeseries. This dataset can also serve as a positive example pool for users with known exposure to ILI when joined with datasets of healthy user's biomarkers for classification of ILI.

4.5 Conclusion

In this chapter we address **RQ.2**: How sensitive are the biomarker signals sensed by commodity smartwatch fitness trackers to the earliest stages of ILI infection and can these signals be used to detect infection pre-symptomatically or even for asymptomatic cases? We do this through a first-of-its-kind flu challenge study with commodity smartwatches that are already adopted across society at scale.

This work shows that significant changes are observed in digital biomarkers captured from these commodity wearables before onset of human-perceptible symptoms or accumulation of measurable viral load via PCR tests. This indicates the utility and quality of biomarkers like resting heart rate, heart rate variability during sleep, skin temperature, and skin conductance for early detection of influenza, even for asymptomatic and mild cases. This is particularly important for mild and asymptomatic carrier cases which do not present traditional symptoms like fever used as indicators to self-diagnose and would otherwise go undiagnosed by traditional measures. Digital biomarkers can therefore serve as complementary indicators of flu infection alongside symptoms and PCR tests.

Chapter 5

ProxiCycle: Passively Mapping Cyclist Safety Using Smart Handlebars for Near-Miss Detection

In this chapter, we discuss the development of a small, affordable, and convenient to install device which can retrofit any bike handlebars to introduce proximity sensing for detecting close-passes. We then demonstrate the performance of this system in an in-the-wild pilot study across 7 users, and deploy it in a 2 month longitudinal study with 15 users across 1,000 total miles traveled. We then show that this crowdsource measure of close-passes can be used as a surrogate signal indicating varying cyclist safety across the road network. Specifically we compare this geospatial metric against existing locations of historic car-to-cyclist as well as a crowdsourced measure of cyclist perceived safety. This work was published and presented at ACM CHI 2025 – conference on Human Factors in Computing Systems, the premier international conference of Human-Computer Interaction.

5.1 Introduction

Cycling is a form of *active transport* which offers well known health and environmental benefits. In many industrialized nations the vast majority of adults do not meet physical activity

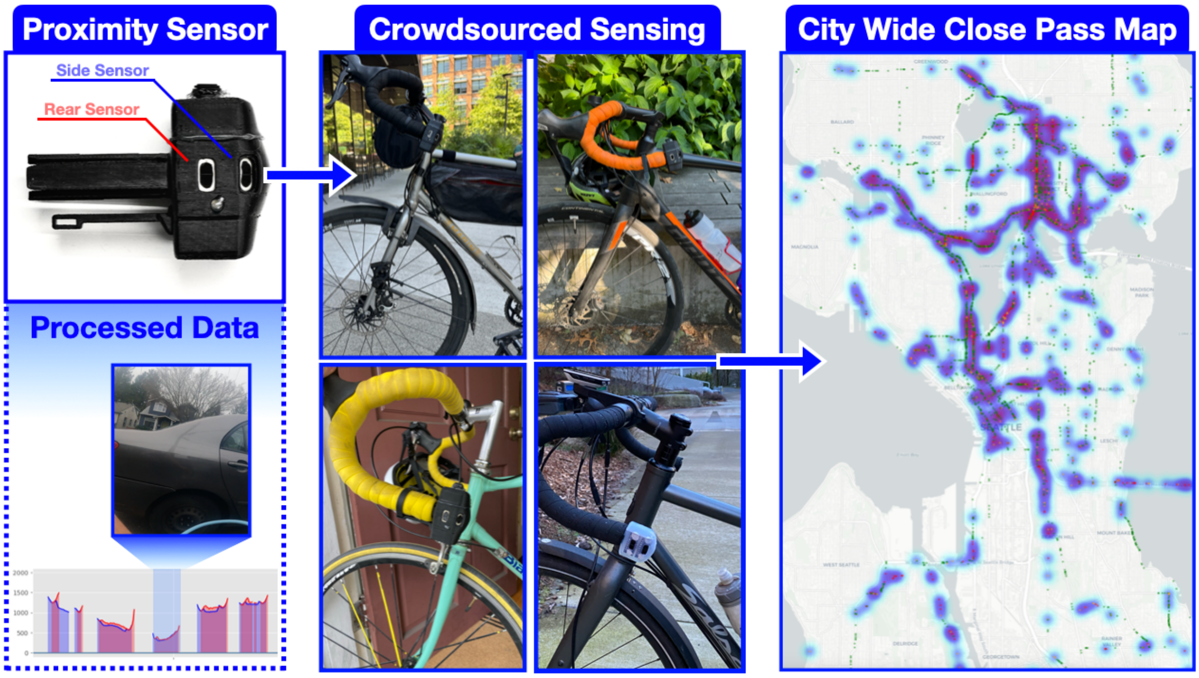

Figure 5.1: ProxiCycle leverages a dual proximity sensor design to retrofit and empower existing bike handlebars with the capacity to passively capture distance of passing cars during overtaking events. This data, when aggregated across many users, can provide useful insight into the safety of the road network without depending on users self-reporting.

guidelines to prevent chronic and fatal diseases such as cardiovascular disease or metabolic syndrome [134] which are responsible for the majority of mortality worldwide (approximately 35% of global fatalities [9]). Regular cycling and other forms of active transport have been shown to lower risk for cardiovascular disease, obesity, diabetes, some cancers, and even psychological disorders, demonstrating just some of the many health benefits [135, 136, 134].

Cycling as a mode of transport also has similarly positive impacts on environmental sustainability. For example, if the global population leveraged cycling for 2.6 kilometers (1.6 miles) a day (the current average in the Netherlands), global emissions from passenger vehicles would drop by 20% [137]. This is significant as transportation by personal passenger car is repeatedly found to be the largest single contributor to the transportation sector's greenhouse gas (GHG) emissions [138] with the transportation sector being the largest portion of total GHG emissions worldwide [139, 138]. This motivates mode switching from car-based to cycling (or other active transit like walking or public transport) as one of the most direct approaches an individual can take to better their own health and reduce their environmental impact. The United Nations highlights this in goal 11 of "17 Sustainable Development Goals to combat climate change by 2030", which states improving the safety and broadening access to cycling (especially for vulnerable users such as women, children, and people with disabilities) as a primary contributor to sustainability [140, 141].

In order for populations to reap the economic, environmental, and health benefits of cycling, the majority of people need to feel safe doing it. Unfortunately, many people do not make this switch. For example, in the US alone, cycling trips make up less than 1% of all daily trips (for reference the leading country in cycling, the Netherlands, sees 25% of all trips including errands and commuting made by bicycle and 22% made on foot) [142]. Investigation into factors preventing cycling amongst Americans reveal that perceived safety, that is, how safe individuals feels irrespective of empirical measures of safety, is the main deterrent to the adoption of cycling, and specifically, the fear of injury or death due to collisions with cars [143, 141]. This belief is not unfounded as every year $\approx$ 41,000 cyclists die from road traffic crashes worldwide [141]. This number is disproportionately large when compared to total cycling trips. For example, in the UK in 2019, 6% of all road fatalities were people cycling despite cycling representing just 1% of all distance travelled and 2% of trips made.

Growing the cycling population is most directly achieved through modifying roadways to improve safety by separate cyclists from automobiles. For example, in Seville (Spain) the cycling population grew by 435% in one year – the same year the city expanded the protected bike lane network from 12 kilometers (7.5 miles) to 151 kilometers (94 miles) [144]. The addition of bicycle infrastructure indicates increased safety, which has been shown to be the deciding factor for many people adopting cycling [145]. However, changing the built environment takes time and political investment. So what about routes which do not yet have dedicated bike infrastructure? While more structural change is needed to make existing routes safer, one way cyclists can have immediately control over the safety of their cycling experience is to ride on known-to-be safer roads. Many experienced cyclists already do this is by developing their own preferred routes, defined by Kevin Lynch as *mental-maps* [82, 146] , by learning safety cues through trial and error. However, this approach is highly personal and developed through prolonged exposure to safety risks from riding on unsafe roads – something that many novice cyclists are not willing to do (i.e., explore-exploit dilemma). Instead of asking every cyclist to take on this up-front risk, a more accessible approach could be to aggregate existing knowledge of perceived safety captured by experienced cyclists and systematically communicate it to novice cyclists to help navigate them on safer and more comfortable routes as soon as their first ride. This can both mitigate safety risk by directing cyclists on safer routes and can lower the barrier of entry to many new cyclists, which can have further compounding effects on safety due to the principle of Safety in Numbers [147]. This measure of perceived safety, also referred to as level of stress (LOS) or level of traffic stress (LTS), [148] has been shown to influence people's decision to cycle over other modes, as well as increase comfort while riding [145, 149]. The perception of safety has a particularly strong impact on cycling adoption amongst parents and children [150, 151]. Prior work has called out a need for better quantitative measures of cycling perceived safety directly as a primary approach to promote cycling [152, 153, 154, 155].

However, existing systems for mapping safety and perceived safety are limited. For example, safety is measured primarily through historic bike collisions which are sparse and slow to reflect infrastructure changes [156]. Safety navigation therefore often defaults to simply reflecting the existence of bicycle infrastructure or lane width [157], lacking any further insight into other factors influencing safety or cyclist stress [82]. Observational studies or cyclist surveys aim to improve data quality [155], but like most survey methods, suffers from low coverage and high degree of aggregation [154, 155]. Platforms like Strava [158] have improved representativeness and granularity of cycling traffic data by involving citizens in open cycling data collection, but this paradigm has not yet been applied to safety [159]. Traditional car-based road safety literature addresses this using *collision surrogates* which are measurable events which are not, but correlate to, collisions such as *near misses* [34], but currently no scalable analog exists for bikes. Such a measure could serve as an indicator of collisions and therefore an indicator of empirical cycling safety. The current state of the art for sourcing bicycle near miss data is manually labeled video footage from constantly recording cameras such as the 100 Cyclist Project [160] or self-reporting through ad hoc web-portals like UpRide [78]. While these systems promise to leverage the scale of crowdsourced data to identify hotspots, they rely on manual reporting which does not scale. This is on top of the fact that cycling collisions themselves are already under-reported [161]. Not only is this tedious for users, but it is also prone to user bias as cameras do not capture true or standardized distance of near misses. Despite their flaws, the existence of these platforms and their user-bases motivates a more scalable bike safety sensing system.

To this end, we investigate a commonly reported unsafe (and uncomfortable) event in which a car overtakes a cyclist too close – a so-called “close pass”. A close pass is when drivers overtake a cyclist on the road dangerously close, nearly colliding with or side-swiping the cyclist. In the field of injury prevention, Heinrich’s Triangle of Injury Prevention proposes that these near-collisions are related to true collisions and avoiding them will in turn lead to a reduction in more severe collisions [162]. Additionally, these events influence cyclists’ perceived safety and in turn their willingness to ride [148, 163]. Many countries including the US, Germany, Belgium, and soon Japan [164, 165, 166, 167, 168] have legal distance margins that cars must provide when overtaking a cyclist, further grounding our definition of a close pass as a signal for safety. Over half of all US states have laws prohibiting drivers from passing cyclists within 1 meter (3 feet) as a safety measure, with as much as 1.5-2 meters common in some European countries [167, 164]. Despite this well-defined guideline, there is currently no scalable method for sensing and reporting where, when, and how severely drivers break these guidelines. To address this gap, we investigate the use of a bike mounted proximity sensor to automatically identify these close pass events passively while cyclists ride to provide a standardized, quantitative, and input-free method for sourcing a measure of cycling safety across the road network. The contributions of this paper are as follows:

1. A formative study via a survey of cyclists of varying expertise to understand the potential efficacy of empirical safety data in facilitating cycling.
2. The design of a smart proximity sensing handlebar plug and technical validation demonstrating the feasibility of passively measuring the distance of passing cars as cyclists ride, and
3. A 2-month longitudinal deployment with 15 local cyclists to develop an aggregate map of observed close passes as a measure of cyclist safety which we then validate against historic collision data and perceived safety surveys.

ProxiCycle advances the line of inquiry on bicycle safety across the road network, motivated by the current lack existing approaches for measuring bicycle safety in a standardized manner at scale. We draw on traditional automotive collision surrogate literature [34] and existing bicycle crowdsourcing systems [169, 158, 170] when designing ProxiCycle to be scalable, low-cost, fully passive, and easy to deploy. Our system contributes to the growing field of smart active transportation and bicycle HCI [171, 172, 94, 173, 95, 174, 175, 176, 152].

5.2 Formative Study

We conducted a brief formative study through a locally distributed fully anonymous IRB approved survey to understand the potential affordances of a map of cycling safety. This survey was designed to answer the following questions: (1) what do cyclists consider when making the choice of whether and where to cycle, (2) how might a map of perceived safety impact these choices, and (3) do answers to these questions differ across cyclists of varying experience?

5.2.1 Survey Structure

We recruited survey participants through advertisement within the university, local neighborhood newsletters, social media messaging, and snowball sampling. After 3 weeks we received 389 responses. Participants were asked demographics questions and then to self-report their experience with cycling on a scale of 1 = very inexperienced to 7 = very experienced (participants were prompted that experience was defined as combination of frequency and consistency cycling). We then asked a series of five questions probing their perception of the safety of cycling, what factors influence their perception, and whether they would leverage knowledge of safety across the road network to influence their choice to cycle more. A full list of survey questions as they appeared in the survey is included in the Appendix **??**. To prioritize recruiting a large survey population, we designed the survey to be completed quickly utilizing a combination of multiple choice and Likert scale questions. Participants were also given an optional short answer response for each question to further elaborate on their answer.

5.2.2 Survey Responses

Of the 389 responses, 71% self-reported above average experience cycling (experience 5-7 out of 7) indicating our sample population tends towards experienced cyclists. We discuss responses to each question in turn below. In the following discussion we provide quotes from the open response questions along with the survey participant ID and their self-reported cycling experience rating.

Perceived Danger of Cycling: The mean response was 3.61 out of 5 with less experienced

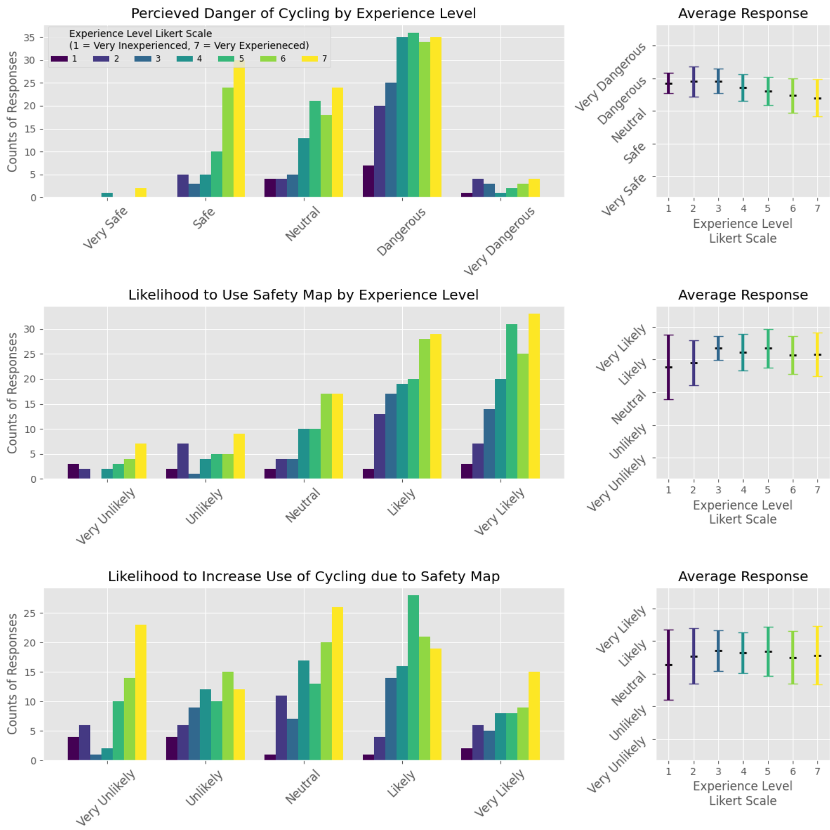

Figure 5.2: Three grouped histograms showing the count of responses for the three core survey questions, grouped by the self-reported experience level of respondent on a 7-point Likert scale (from 1 = very inexperienced to 7 = very experienced). The mean and standard deviation for each question by experience level are shown to the right. The histograms show, from top to bottom, the respondent's perceived danger of cycling, their likelihood to use a mapping application which communicates variable safety across the road network, and their likelihood to increase their utilization of cycling for transportation as a result of this mapping tool.

riders (experience 1-3 out of 7) ranking cycling as more dangerous than more experienced cyclists (experience 4-7 out of 7). Less experienced riders (experience 3 out of 7) ranking it the highest at 3.91. These results are illustrated in Figure 5.2. About a quarter of responses used the open response section to elaborate that cycling itself is not dangerous but rather having to share the road with cars is. One participant said *"the danger is outside of my control. My biking is safe, it's others that make it dangerous."* (S27 6/7). Other respondents shared similar sentiments that: *"The danger from cycling comes from cars."* (S5 6/7), and *"Cycling as a mode of transit is very safe. Traffic is not safe but that is not cycling's problem"* (S165 7/7). Some participants even brought up the varying safety across the road network as a result of road-specific infrastructure and driver behavior – a key motivation for this work. One participant said *"depends on route/infrastructure, and how accustomed drivers are to being around bikes"* (S35 7/7), while another mentioned *"I work very hard to avoid high traffic volumes and*

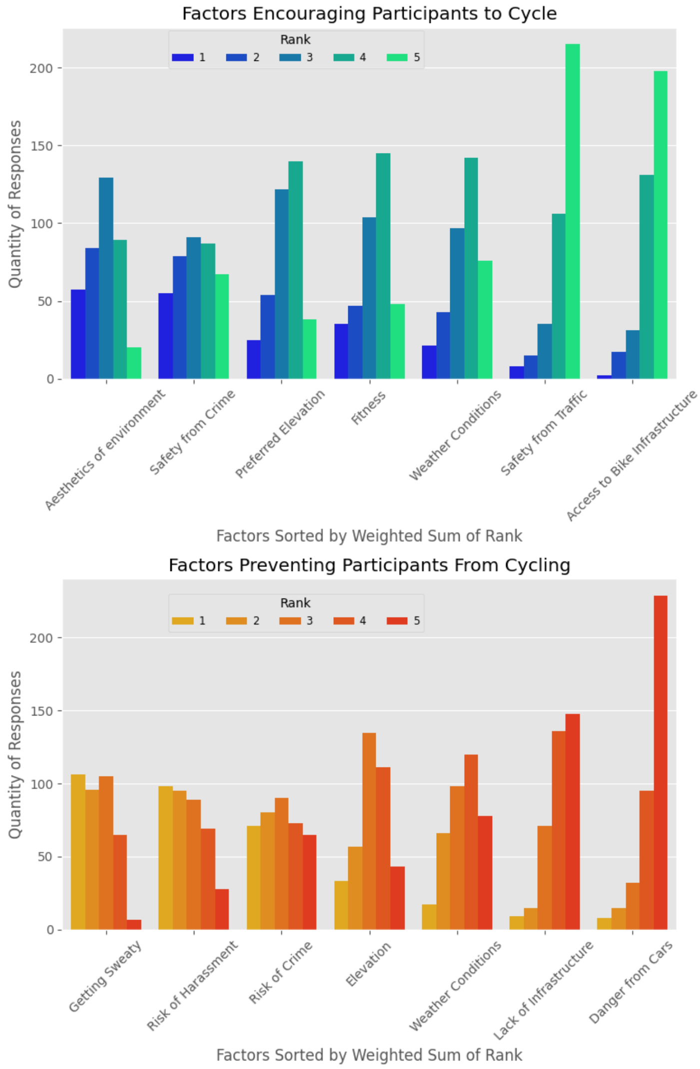

Figure 5.3: Two grouped histograms showing the ranking of influence (from 1 being lowest to 5 being highest) for each factor encouraging (top) and discouraging (bottom) respondents from cycling. Each histogram shows the total number of respondents who assigned each rank to each factor. The factors are sorted in order of weighted average rank from lowest (left) to highest (right) influence.

high speed [roads], and because of that I feel mostly safe. During most trips, however, I have to face at least one unsafe segment or intersection." (S122 7/7). Some participants went further and noted that knowledge of varying safety across the road network can help improve their experience cycling. One notes, *"Knowledge of streets and traffic patterns is essential, but once you have a familiar and well considered route, it is very safe"* (S245 7/7), and another says *"You do have to know which streets to ride on and ride defensively"* (S95 7/7). Another participant highlights the frustration of feeling forced to do this legwork, motivating a platform to share this information across the cycling population: *"The single worst part about cycling here is [those who ride] bike or walk either have to research their route or risk being thrown* *into active traffic."* (S223 3/7) Participants like this directly cite the barrier of entry to novice cyclists. These results indicate that less experienced cyclists perceive cycling as more dangerous and affirm our expectation that experienced cyclists may ride on a safer road network by leveraging prior knowledge to avoid dangerous streets (i.e., mental maps accumulated through experience) – something that we aim to share with the broader cycling population.

Ranked factors encouraging and discouraging cycling: Next, participants were asked to rank factors which encourage or discourage cycling. The ranking stratified by respondent's cycling experience is shown in Figure 5.3. Among the factors which encourage cycling, access to infrastructure and safety from traffic were ranked most highly. Among factors which prevent cycling, danger from cars was ranked most highly by a significant margin across all experience levels. A few participants cited personal experiences with close-calls or collisions which have since detured them from cycling *"having had plenty of close calls with inattentive drivers I have little interest in riding on roads, or even unprotected bike lanes"* (S340 4/7). This result indicates that conflict with cars is a significant factor preventing cyclist adoption.

Likelihood to use safety map and perceived effect on cycling: Respondents indicated they would be very likely to use a safety based navigation map with an average response of 4.36 out of 5. Those with intermediate levels of cycling (3-5 out of 7) were most likely to use it. These results are illustrated in the top row of Figure 5.2. Finally, when asked how likely they were to increase their frequency of cycling given access to this data, Figure 5.2 shows respondents overall were more likely to increase their cycling with a mean response of 3.62 out of 5. The mean response was greatest for below average experience (3 out of 7) cyclists at 3.73 which intuitively makes sense as the most experienced cyclists may already cycle at max capacity or feel secure with their "mental map" of safety given their increased experience [82]. This intuition is reinforced by short-answer responses from expert cyclists (7 out of 7) participants who note, *"A map will not help me cycle more. I already use my own senses and memory and find the best routes, even for places I've never been"* (S44 7/7). On the other end, those with no experience at all may not have interest in cycling. Though some inexperienced cyclists may want to cycle more but feel blocked by the complexity of route planning: *"Confidence in navigating safe cycling routes is my #1 barrier to cycling more. I don't feel confident that Google Maps or other navigation systems would provide a safe route and despite my relatively high awareness of the Seattle bike lane system, there is still a prohibitive amount of route* *planning research for casual use"* (S365 3/7). Quotes like these demonstrate how experienced cyclists tend to identify these safer routes through trial and error, however, this may not be an option for inexperienced cyclists or even some experts who are not willing to make the safety trade off by cycling unfamiliar territory [82]. This is directly called out by a novice cyclists (1 out of 7) who says *"My iPhone and the Maps app fundamentally changed riding transit for me, an always up to date bike map with honest info would be amazing"* (S146 1/7). That being said, some expert cyclists noted this may still be useful in unfamiliar environments *"I know my way around my home well, but a map of that type would help in places I don't know as well or am new too"* (S139 6/7). These findings further motivate ProxiCycle by identifying that access to this quantified measure of safety is useful to cyclists of all levels, but specifically can make cycling more approachable by increasing the likelihood of cycling among less experienced riders.

5.3 System Design

In this section we will discuss our design considerations when building ProxiCycle as well as the physical and software implementation. We finish this section with a brief road map of the study design from technical evaluation, to deployment, and finally spatial analysis of the crowdsourced map of close passes.

5.3.1 Design Consideration for a Collision Surrogate Sensing System

Our goal with this work is to provide a scalable measure of safety across the road network which can be used to navigate cyclists on safer routes, in turn lowering the barrier of entry to cycling and growing the cycling population. Following traditional injury prevention and collision surrogates research [162], we designed ProxiCycle to measure safety passively as users ride through the lens of close passes as a potential collision surrogate. A strong collision surrogate will occur much more frequently than actual collisions and represent not just the most severe collisions, but all possible collisions, including those which may go unreported. This approach addresses the scarcity of safety data as close passes occur much more frequently than self-reported near misses and collisions and eliminates both user burden and user bias [177].

This decision to use close passes as a collision surrogate is supported by existing work using lane width as an indicator of LTS by urban planners as well as legal passing margins referenced in Section 5.1 [178, 148]. We made this decision despite right-hook collisions at intersections making up the majority of reported bicycle collisions [33] as studies show that severe under-reporting has introduced strong bias into the types of collisions which are available for analysis in favor or the most severe (i.e., those resulting a visit to the emergency room) [161]. Conversely, a naturalistic observation study of risk-factors of road cycling found that *side-swipes* (where a car collides with a cyclist side-to-side while passing) were the most commonly observed hazard, more than twice as common as *right-hooks* (the most commonly reported collision, where a car performs a right turn into the cyclists path) and over 10x more common than *dooring* (where a parked car door is opened in the path of a cyclist) [36]. Finally, prior work indicates that ride quality is improved when riders are most present and less distracted by technology [179], so we constrain our design to: (1) work entirely automatically with no user input, and (2) for our device to be as physically unobtrusive as possible.

5.3.2 Proximity Sensing Technology

The most popular off-the-shelf proximity sensors include: acoustic (sonar), radar, stereoscopic video, and light based time-of-flight (ToF) sensors. Single lens cameras are sometimes used to estimate distance but these approaches are error prone. In this study we chose to leverage the ST VL53L8 multizone ToF sensor [37] for its small size, interpretable output signal, minimal data overhead (on the order of kilobytes per minute) and lack of multipath issues (common with both radar and acoustic ranging). The narrow field-of-view of ToF sensors allows for the system to easily isolate passing cars to the left of the cyclist without interference from other objects in traffic (i.e., parked cars to the right or vehicles traveling in front of or behind cyclists).

The VL53L8 works by sending an infrared light pulse (Tx) in a single direction and listening for a response (Rx) due to reflections off surfaces in range and in the field-of-view. Unlike many ToF sensors, the VL53L8 is specifically designed for outdoor use making it more robust to infrared interference from the sun. ToF sensors are also not commonly used for automobile proximity sensors so there is low risk for interference from other vehicles on the road. Anecdo-

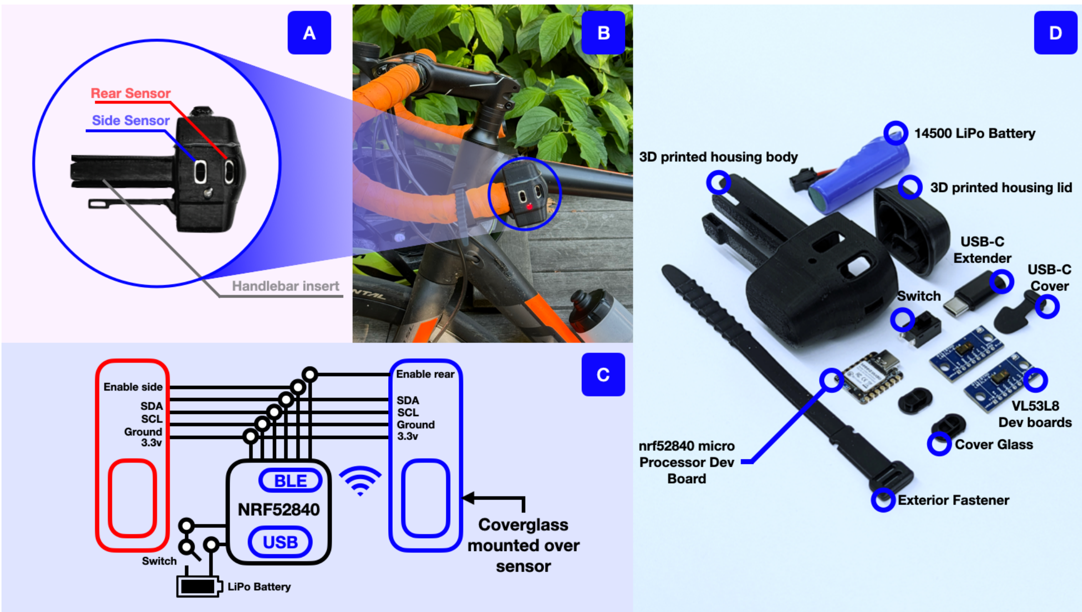

Figure 5.4: A) Close up of the fully fabricated Smart Handlebar Plug. B) The placement of the device on the end of a bicycle handlebar. C) A circuit diagram illustrating how components are connected. D) A layout of components to fabricate the device.

tally, we experimented with ultrasonic acoustic ranging sensors and found they suffered from interference from existing ultrasonic noise on the road. This was likely due to sensors built into nearby vehicles for blind spot detection. The small size of data produced by the VL53L8 also allows for easy communication over I2C and BLE, even at high data rates. These ToF sensors are also smaller than parallax based depth camera alternatives, allowing for multiple sensors to be integrated onto the same device without compromising the unobtrusive form factor and do not suffer from self-interference like sonar and radar might. Finally, the data from these sensors do not produce visual features and are inherently privacy preserving.

5.3.3 Hardware Design & Placement

Physical Placement: When designing our custom sensing platform we constrained our sensor's physical design to: (1) preserve limited handlebar real estate, and (2) position the sensor outside the rider's grip position to avoid measurement noise from the rider's hand or arm blocking the sensor. This led us to the final design of a smart handlebar plug by integrating the proximity sensor into the plug used to secure handlebar tape at the end of the handlebar. While the current prototype is designed for drop bar handlebars, similar designs can exist for flat bars or other handlebar shapes so long as the designs maintain an unobstructed field-of-view.

Our sensor is mounted by sliding it into the open end of the hollow handle bars. This design can be seen in sub-figure A) and B) of Figure 5.4. Given the small size of these sensors, we were able to integrate both a side facing sensor and a rear facing sensor to understand the velocity of passing objects based on the order of sensor triggers. This design avoids consuming handlebar real estate, leaving room for other devices such as bike computers, phones, or bells to be mounted on the handlebars[1] while also placing the sensor as far to the left of the bike as possible, which simultaneously increases the sensor range with respect to the bike and minimizes the likelihood of the rider's arms interfering with the sensor's field-of-view. The whole device including 3D printed housing is 6cm by 4cm by 3cm and was primarily constrained by the 14500 battery used for this prototype. However, we did not prioritize size optimization in this work and the majority of the housing is empty space to accommodate future development with additional components. The components themselves are 2.7cm by 2.7cm by 0.9cm and can be further reduced using a custom PCB.

Hardware Components: The sensor data is captured from 2 VL53L8 proximity sensors and communicated to a generic microprocessor (NRF52840) on the Seeed Studio Xiao development board over I2C. The Xiao comes complete with a Bluetooth Low-Energy (BLE) antenna used to transmit the data to a paired smartphone. The system is powered by a 3.7v 1000mAh 14500 LiPo cell battery to maximize battery life while continuously collecting raw data. The components of the device and block circuit diagram can be seen in sub-figures C) and D) of Figure 5.4. Each of the VL53L8 sensors in continuous mode have a current draw of 40mAh [37] and thus a 1,000mAh battery will last over 12 hours powering both sensors in ideal conditions. This is a reasonable battery life as cyclists often use bike lights which typically have $<$ 8 hours battery life. The battery capacity, and therefore size, may be significantly reduced once power optimized for on-device processing and limiting wireless data transmission. The Xiao also natively supports power over USB-C allowing the device to be powered by external battery packs, dynamo motors, e-bike charging ports, or even the smartphone itself. The whole system can cost less than $25 to build in house using off-the-shelf components which is already an order of magnitude cheaper than the closest alternative (i.e., Garmin Varia $200).

[1] This proved significant in our user-study as the majority of our participants had fully instrumented their handlebars leaving no room for additional attachments. This design may therefore improve the usability of the device.

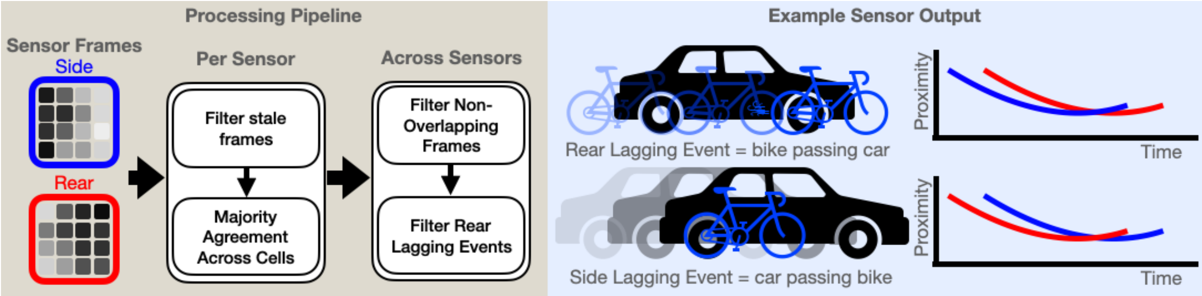

Figure 5.5: Flow diagram of our signal processing procedure for data recorded using our custom sensor platform for close pass event recognition. At each frame of each of the two sensors the system checks if the majority of cells updated agree within some threshold τ of each other (i.e., most cells update to a similar value at the same time step) to identify sensor events. Then the system checks if there is overlap between sensor events from each sensors and logs a close pass if the rear sensor event triggers prior to a side sensor event indicating that the passing object has a positive relative velocity with respect to the bike.

The fabrication cost (i.e., PCB, plastic housing, and assembly) could be further reduced with manufacturing at scale.

5.3.4 Signal Processing

To process the readings we employ a signal processing chain shown in Figure 5.5. The raw data from the VL53L8 sensors are low-resolution 4x4 frames representing the depth at each of the 16 cells within the sensor's field-of-view at the time a frame was captured. Each sensor samples at 10Hz including BLE transmission delay. To convert this raw timeseries data into meaningful sensor events, we first filter out any cells unchanged from the previous frame (which occurs when no objects are in range for that cell). We then log a sensor events if the majority of cells in the sensor frame agree to the same value within a threshold parameter, which we set as 0.2 meter empirically through early experimentation during development. This ensures that sensor events are only logged when large objects like cars block the sensor's field-of-view , i.e., one or more small objects like branches or distance pedestrians in the sensor's field-of-view will not trigger this condition. While there is no guarantee that this approach strictly identifies cars, we empirically saw during testing that passing cyclists and pedestrians rarely satisfied this condition at the typical passing distance due to the wide field-of-view of the VL53L8. The 0.2 meter threshold was used to give a small tolerance to natural noise between cell's measured distances or reflections off surfaces with modest curves while still maintaining agreement between cells. We then log events as a passing event if both the

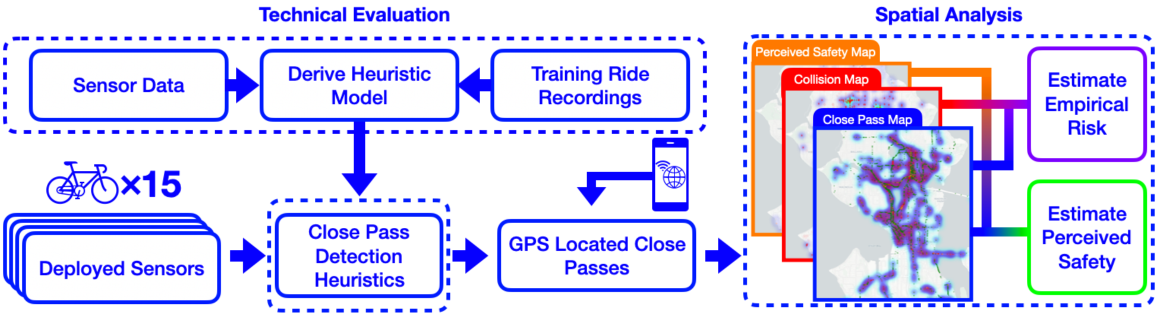

Figure 5.6: An overview of the study road map. The study begins with a technical elevation to derive the signal processing technique using raw sensor data compared to video recordings. The we leverage this model during a longitudinal sensor deployment to source GPS located close passes which are used in a spatial analysis of bicycling safety across the traveled routes.

rear and side sensors trigger events overlapping in time and the rear sensor is triggered first. This lag in the side sensor indicates that an object passing the bike approached from the rear with a positive relative velocity with respect to the bike. Events where the rear sensor lags behind the side sensor indicate a negative relative velocity, i.e., cyclist passing a static object or slower road user. Since cars can pass cyclists with variable velocity, we did not threshold the magnitude of the lag between sensors. This ensures that cars would be detected, irrespective of their speed. This relationship is illustrated in Figure 5.5. We then compress the 16 cell ToF sensor timeseries signature to a single pass distance using the inner-quantile mean (IQM) of the side sensor at each frame and select the minimum IQM across the duration of the pass. We then filter out any pass distance above that of a large road lane at 3 meters and below 0.2 meters as these short distance triggers are due to cyclist hands during grip adjustments. We use the legal definition of a close pass as 1 meter and consider sensor events at the 1-3 meter range to be safe passes. In practice, we identified borderline cases where 1.3 meters felt uncomfortable to our participants, and for the purposes of evaluation, we report close pass and borderline close passes (<1.3 meters) as positive. This is closer to the most progressive legal passing margin in the US at roughly 1.2 meters (4 feet) which was written into law the same year this paper was written [180].

5.3.5 Study Road Map

Figure 5.6 shows a road map of ProxiCycle evaluation. After completing hardware development, we conduct a technical evaluation in Section 5.4 to validate the signal processing discussed above in Section 5.3.4. We then conduct a longitudinal deployment which we discuss in

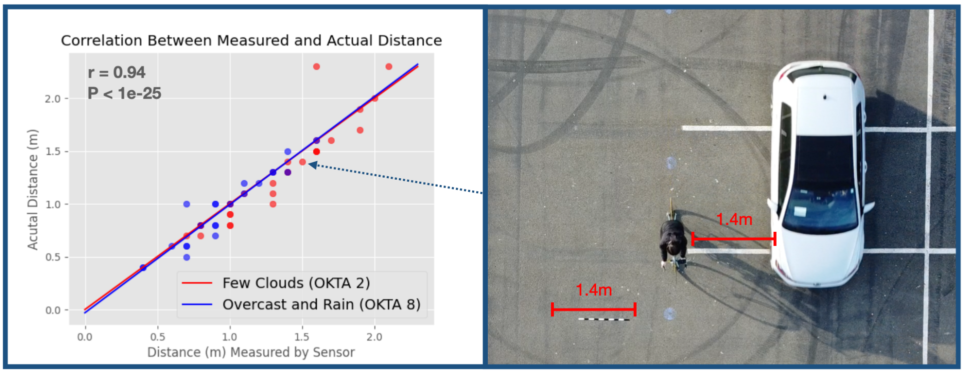

Figure 5.7: (Left) The correlation (Pearson's r of 0.94 and P-value < 1e-25) between ground truth distance measured manually in pixels from aerial view using reference marker and ProxiCycle 's distance readings captured on both sunny and rainy days. (Right) A frame from a drone video illustrating how ground truth distance was captured using the reference marker in frame.

Section 5.5 to crowdsource a city-scale dataset of close passes. We finally show that this map of close passes can be used to analyze safety across the road network in Section 5.5.1.

5.4 Technical Evaluation

In this section we conduct two technical evaluations. First, we validate the chosen sensor modality is capable of measuring proximity accurately in the specific handlebar position and context of both moving bike and moving cars in multiple weather and ambient lighting conditions. This validation is essential to ensure trustworthy readings during in-the-wild deployments where it is not feasible to collect ground truth distance. Next, we investigate potential sources of error in close pass detection through a 7-user IRB approved pilot study where we compare triggered sensor events against manually reviewed simultaneously captured video footage on unrestricted routes.

5.4.1 Experiment 1: In Lab Validation

As a proof-of-concept we investigate the accuracy of the VL53L8 when mounted on the left stem of the bike handlebars[2] for sensing the proximity of passing cars through a controlled in-lab experiment. To ensure the results of this experiment most directly reflect the real world

[2]We note that all user study trials were conducted on roads with right lane driving where bike lanes were always right of the nearest traffic lane. We note that this set up fails to capture close passes to the right of the cyclist. This system can easily be extend to capture both sides by adding a second device.

interaction, we chose to simulate passes with both a moving bicycle and moving car.

We used photogrammetry (i.e., simultaneous video capture) to capture a reliably ground truth distance to compare to our sensor readings. To do this, we leveraged a 1 meter stick painted alternating black and white every 10 centimeters in frame as a visual reference of pixel count to distance in our video. We found it was easiest to consistently capture space between the bicycle and car from an aerial perspective (as opposed to a tripod on the ground in-line with the pass). So, we recorded video from a DJI mini 2 drone in hover mode directly above to simulated passes. By capturing video of both the simulated pass and black and white reference marker, we were able to manually map pixel quantity of each 10cm subsection of the marker to the real world distance at the 10cm resolution. Figure 5.7 shows the length of the reference marker compared to the length of a pass in the same frame. Because there was potential for the altitude of the drone to shift during flight, the mapping of pixel to distance was recalculated for each pass (though in practice it remained relatively constant). A sync gesture visible from both the camera and the sensor was performed at the start of each data collection.

By flying the drone at a sufficient altitude and centering the location of simulated close passes in frame, we can minimize any possible effects from edge distortion (e.g. barrel distortion) due to curvature of the lens. It is not essential to use a drone, as a similar approach could be replicated using a camera mounted from a tall structure such as a building given proper orientation and line-of-sight. For all trials both the cyclist and car were moving in a straight line to simulate the most common approach of automobiles overtaking a cyclist. the car was traveling between 24 and 48 kilometers per hour. To assess the effects of weather (both lighting conditions and precipitation) on sensor readings, we repeated the same experimental procedure twice: once on a mostly sunny day with few clouds and again on an overcast rainy day. The cloud cover on these days were recorded using OKTA scale, a meteorological standard metric, and were recorded as OKTA 2 and 8 respectively.

We captured 42 simulated rectilinear passes at variable distances from 0.4 meters to 2.2 meters and found our sensor readings to have high correlation with actual distances with an aggregate Pearson's r of 0.94 (0.93 in sun and 0.95 in rain). This relationship is statistically significant with an aggregate P-value of 1e-25 (1e-13 in sun and 1e-11 in rain) indicating that the proximity metric captured from our smart handlebar plug is highly correlated with the

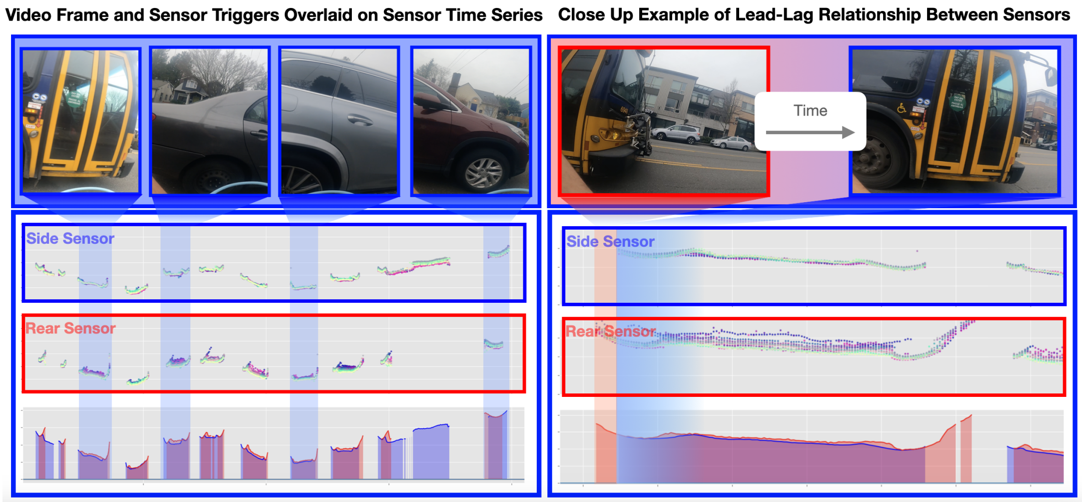

Figure 5.8: (Left) A visualization of the video capture overlaid with processed sensor triggers. (Right) A close up example of the lead-lag relationship used to distinguish passing vehicles from instances of the cyclist passing something in the environment.

ground truth distance with no significant variation due to lighting conditions or precipitation. These results can be seen in Figure 5.7 alongside a example frame from our ground truth recordings.

5.4.2 Experiment 2: In the Wild Pilot Study

Next we deployed our system in-the-wild with a group of 7 pilot riders (each riding their own personal bicycle) in Seattle to evaluate the detection accuracy of our system at identifying cars against other real-world confounds. We deployed each sensor module in conjunction with a GoPro action camera mounted on the handlebars pointed perpendicular (left) to the bicycle to record passing objects during each user's ride. In the video we capture all instances of passing objects including cars and other large objects which trigger our sensing system. A sync gesture visible to both the sensor and camera was performed at the start of each ride. The synchronized video and sensor feed were manually reviewed and annotated for evaluation.

Our participants were recruited through social media messaging and word of mouth from a local bike commuting community. All participants road their own bicycle retrofit with our sensor for the study. Self-reported experience level of our riders varied with: 4 cycling a few times a month, 2 cycling multiple times a weekly, and 1 who cycled multiple times per day. Each user road with our sensor and GoPro for 45 minutes to 1.5 hours (limited primarily by the GoPro battery life). The GoPro was mounted on the left side of the handlebars facing left (i.e. towards passing traffic). The routes were not pre-defined to preserve the naturalistic discovery of possible sources of error. Riders primarily chose familiar routes during their rides. The rides were carried out on different days with varying weather conditions including light rain, clear skies, and overcast skies. All trials were conducted during daylight hours to ensure the video feed from the GoPro was interpretable.

To generate ground truth labels for evaluating ProxiCycle 's performance, the accompanying video footage was manually annotated.[3] Because of the lack of ambiguity in these labels, i.e., car pass or not, all videos were annotated by a single annotator over multiple days in hour long sessions with breaks in between sessions to avoid fatigue. Annotations were made by scrubbing through the videos one-by-one from start to finish and recording the timestamps where cars passed the cyclist. This was done over two passes. First, an initial blind pass, i.e., without access to ProxiCycle sensor data, was made to visually annotate every time a passenger automobile, i.e., car, truck, bus, etc., passed the cyclist in the footage. Then a second pass of each video was made by the same annotator, this time with access to the sensor events recorded by our signal processing technique defined in Section 5.3.4 to compare ProxiCycle against the annotations from the first pass. Sensor events which were not accompanied by a passing automobile were labeled false positive and passing automobiles unaccompanied by a sensor event were labeled false negative. These error statistics were binned by the IQM of the raw sensor pixels at the time of the event as close (<1 meter), borderline (1-1.3 meters), and far (1.3-2 meters) for later evaluation. Anything outside the 2 meter range was excluded from evaluation due to the context of the problem (greater than the width of road lane) and VL53L8's technical limit (increased error above 2 meters). Automobiles which were greater than 1 full lane of separation in the video were marked as true negative for this reason.

For the purposes of evaluating close pass detection, we treated both close and borderline passes as positive predictions and produced a separate evaluation metric for far passes. By this definition, all users experienced at least one close pass during their ride. This allows us to evaluate the performance of the system within the range of interest of 0 - 1.3 meters while still reporting the evaluation metrics for the full range of the sensors. Interestingly, during post-ride questioning, 3 out of 7 users reported experiencing at least one close pass while they

[3] While this approach could be scaled up by using object detection models for larger evaluations, we opted for a fully manual approach to ensure the highest quality labels in this initial validation.

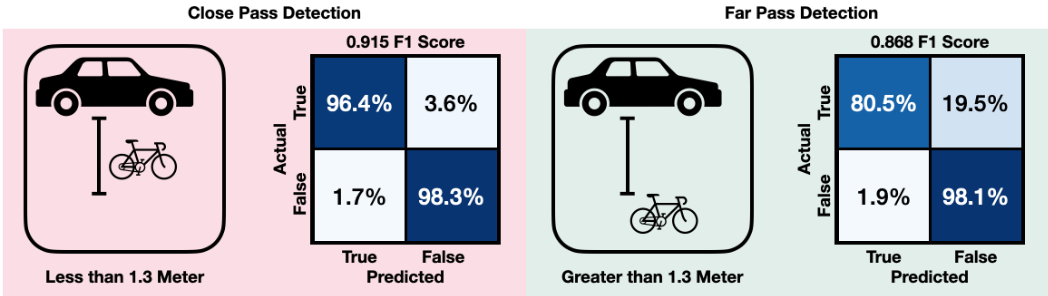

Figure 5.9: The classification results from detected passing objects for both the close pass and borderline close pass region (sub 1.3 meters) and the far pass region (between 1.3 and 2 meters) as well as the categories of common sources of error for both close and far.

actually only experienced borderlines close passes at 1-1.3 meters. This may indicate that the currently accepted legal definition of a close pass may still be too small compared to the tolerance even experienced riders are willing to accept when describing a safe passing distance. The F1-score for close pass detection was found to be 0.915 and the F1-score for far passes was found to be 0.868. These results are illustrated in Figure 5.9.

Of a total of 269 sensor triggers events within the range of interest, we saw only 5 triggers which were not an automobile. While already small, this can be further mitigated through aggregation in a crowdsourcing setting [181]. Additionally, due to the directness of the sensing modality, i.e, physically measuring proximity, these errors can can sometimes be interpreted given context. Qualitative examination of these errors revealed our errors to be primarily at the maximum boundary of the sensor range (close to 2 meters). Interestingly, our processing heuristics were able to correctly filter out most pedestrians, other cyclists, bollards, e-scooters, trees, and even motorcycles traveling both with and opposite the cyclist. This is desirable as these road users are often within the close passing margin but do not pose the same level of risk as passenger automobiles. We found 2 false positives due to static objects to the left of the cyclist while riding on footpaths, 1 false positives due to bollards on protected bicycle lanes, 1 due to a passing cyclist passing very close, and 1 due to the a participant making a very sharp turn causing the ground to trigger the sensors. In all cases, these errors occurred in scenarios which may be identified by including additional sensors. For example, using geographic context from GPS to identify riding on bollard protected bike lane or foot path where cycling near pedestrians and bollards is most likely to occur, and using the smartphone Inertial Measurement Unit (IMU) to filter out data during sharp turns.

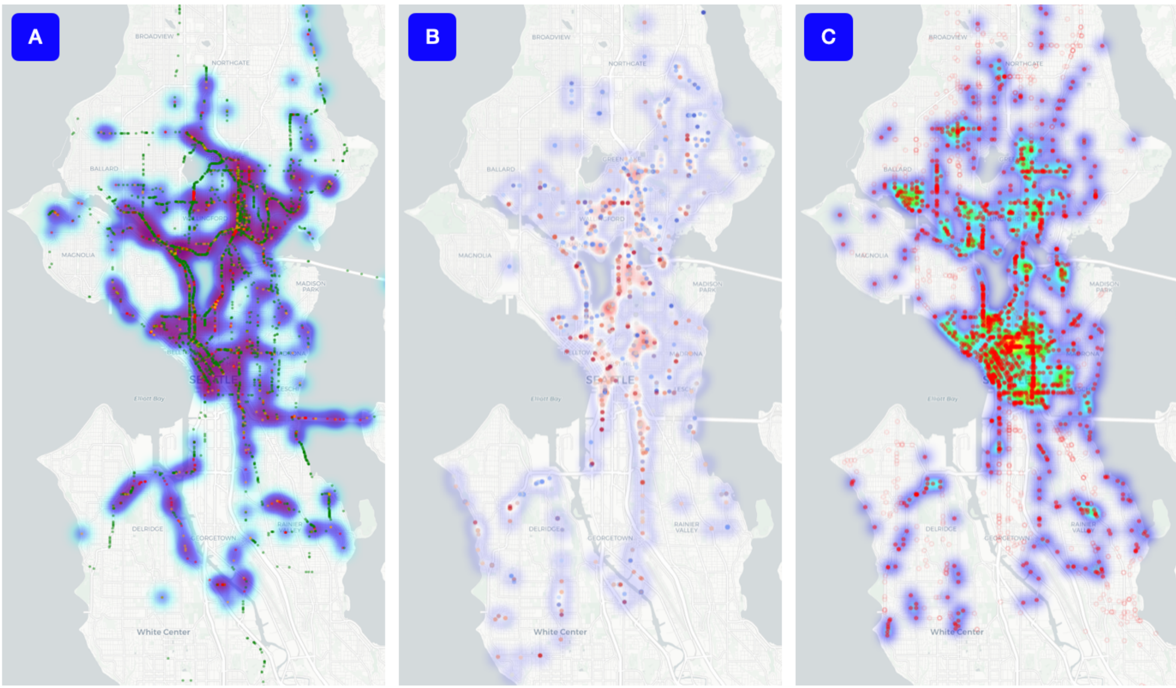

Figure 5.10: A) shows a heat map of close pass density across the routes traveled by participants during our study. Colored points are overlaid at each sensor event trigger including legal far passes, borderline close passes (between 1 meter and 1.3 meters), and close passes (less than 1 meter) in green, orange, and red respectively. B) Shows a heat map and points overlaid of survey respondent's perceived safety rating at locations sampled from our participant's routes. Blue points indicate higher perceived safety while red points indicate higher perceived risk. C) Shows a heat map of historic collisions including at least one bicyclist and one automobile over the last 5 years within a 200 meter radius of our participant's routes. Red points representing collision locations are overlaid, with those outside the range of our participant's route plotted as opaque.

5.5 Longitudinal Deployment Study

With the sensing system validated, we moved on to study the relationship of close passes with existing measures of safety across the road network through a 2 month longitudinal study with 15 cyclists. During this study participants rode as usual but with ProxiCycle installed on their bicycle to crowdsense a city-scale map of close passes which we then used to compare against two existing measures of cyclist safety: (1) a map of historic bicycle collisions provided by the Seattle department of transportation as an empirical measure of safety, and (2) a map of perceived safety which we collected through surveying the general population in Seattle.

Unlike in Section 5.4.2 where we recruited from the general population of cyclists, we prioritized recruiting cyclists who ride more frequently and on highly varied routes in our deployment to maximize coverage across the city. So, we recruited through a local cycling activist community: *Seattle Neighborhood Greenways* via their monthly email newsletter. As part of our screening survey, we asked participants how often they cycled (i.e., daily, multiple times a week, once a week, monthly, or more) and for the consistency of the routes they rode (i.e., on a scale of 1 to 10, with 1 being as varied as possible and 10 being same route every time). After 2 weeks we received 206 responses from potential participants. More than half of our respondents had drop bars on their primary bicycle and about a third owned an Android phone. We built 15 sensor modules and recruited from the subset of drop bar and android users to standardize both our application and physical sensor design. We sorted respondents in order of self-reported ride frequency and route variability and recruited from the top of this list working down. This recruitment process resulted in one participant from the study in Section 5.4.2 participating in the longitudinal deployment as well. Participants were not compensated and participated purely through volunteer interest.

Each participant received a single ProxiCycle device and a brief instruction on how to install the device on their bicycle, connect over bluetooth to their phone, and interact with the paired application to disconnect and confirm data sharing. Data was periodically uploaded to a remote server. Over the next 2 months we recorded 2050 close passes over 1,604 kilometers (1,002 miles) traveled over 240 unique bike rides. The map in Figure 5.10 A) shows a heat map of the density of close and borderline passes recorded across the geographic area covered by these rides. At the end of the deployment, we correlated the locations of these close passes with existing measures of cyclist safety in a spatial analysis.

5.5.1 Spatial Analysis

To understand the relationship between close passes and existing measures of cyclist safety, we correlated the locations of close passes (Figure 5.10 A) with both a survey of people's perceived safety at different points on the road network (Figure 5.10 B) and historic automobile-to-bicycle collisions (Figure 5.10 C). We discuss these evaluations in the following sections.

Close Passes as an Indicator of Collisions

Recall that collision surrogates, like our proposed measure of close passes, should be spatially collocated with historic collisions but occur more frequently to serve as an indicator of safety within a shorter temporal window [177]. To demonstrate the quality of close passes as a collision surrogate, we compare the locations of recorded close passes over our 2 month deployment as a indicator of empirical collision recorded and shared publicly by the local department of transportation in the last 5 years. These spatial signals are shown in Figure 5.10 A) and C), respectively. The intentions of this analysis are not to claim close passes alone serve as a comprehensive model of risk, but that close passes are correlated with existing measures of risk and can be a useful signal for studying and modeling cycling risk in the absence of sufficient or current collision data.

We are only able to analyze locations where we have collected data, so we sampled locations across the routes traveled by our cyclists every 100 meters and assigned them a label of high risk if there was a historic collision within the distance margins within the last 5 years and low risk otherwise. We then use the presence of close passes within the same margin to indicate this risk. We show results of distance margins from 100 to 250 meters in Table 5.1, reporting the results for 200 meters inline. We chose to limit the window of historic collisions to 5 years to avoid including outdated collisions which potentially occurred prior to updated infrastructure, while maintaining a large enough window to represent safety across the entire road network (i.e, while 10 years of collisions holds more statistical power, collisions from 10 years ago may not represent the safety of the current road network). We chose to report the 200 meter distance margin as this encompasses the length of a typical block (i.e., distance between two intersections) in the location of our study. This resolution allows cyclists to plan deliberate detours to avoid collision hotspots. Finer resolutions (shorter than a block) may be too small to afford rerouting around hot spots and would inevitably be aggregated to a coarser resolution to be useful.

We found spatial Pearson correlation of close passes and historic collisions to be 0.6 with $P < 0.0001$. We found a precision-recall area under curve (PR-AUC) of 0.83 and macro-F1 score of 0.8 at estimating the locations of collisions using locations of close passes at a 200 meter distance margin. The results across all distance margins are shown in Table 5.1. These results show a significant correlation between the locations of close passes and collisions, indicating that close passes may be a useful collision surrogate in the absence of collision data. Additionally, close passes occur at a much higher frequency and therefore can provide a more up-to-date view of safety across the current road network in a shorter window of time (2 months vs. 5 years).

Close Passes as an Indicator of Perceived Safety

While less indicative of risk than collisions, people's perceptions of safety is often used in place of collision data to study safety across the road network where collision data is sparse or unavailable [148, 163]. We surveyed people's perceptions of safety at the locations with and without observed close passes in our study to compare close passes as an indicator of perceived safety.

In order to collect a measure of perceived safety to compare against, we conducted a survey study (N=76, 21 with self-reported cycling experience below 4) recruited through snowball sampling from the same activist group as our deployment participants. In the survey, participants were provided Google Street View (GSV) links which visualized locations where close passes had been observed during our 2-month deployment and were asked to rate their perception of safety with respect to traffic at each location on a 7-point Likert scale (from 1 - very safe to 7 - very unsafe). Additionally, we included an equivalent number of locations where no close passes were observed during our deployment to get a balanced sample of perceived safety across both potentially dangerous and potentially safe locations. This resulted in a set of 1027 locations to be rated which were subdivided into 27 mini-surveys containing 50 locations each (half where close passes were observed and half where they were not). Each survey participant provided basic demographic information including age, gender, and experience level cycling (prompted similarly to Section 5.2 that experience is defined as combination of frequency and consistency cycling), along with safety ratings for each location and whether they were familiar with the location or not. Some survey participants completed multiple mini-surveys rating more than 50 locations. All locations were rated by at least 2 raters allowing for an inter-rater agreement to be calculated. We used average of pairwise Cohen's Kappa across all pairs of raters with overlapping locations which we found to be 0.42 which is considered "fair to good"

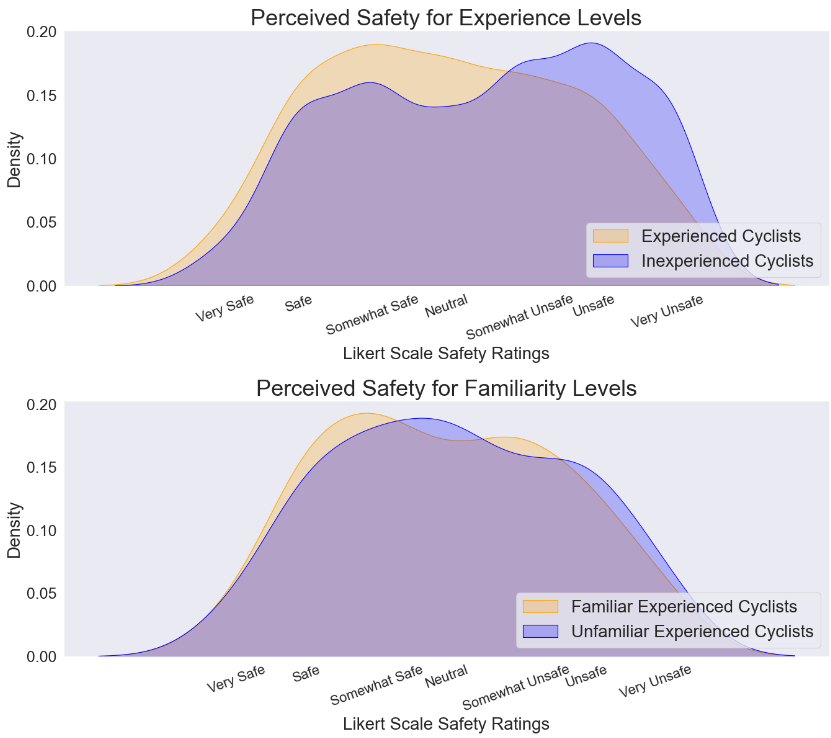

Figure 5.11: (Top) A Kernelized Density Estimate (KDE) showing the distribution of safety ratings assigned by inexperienced (N=21) and experienced (N=55) cyclists. Participants were determined to be experienced by a cut-off applied to self-reported experience of 4 or more on a 7-point likert scale (i.e., participants who self identify as "neutral" to "very experienced"). (Bottom) A similar KDE showing the distributions of safety ratings assigned by experienced cyclists for locations which they were and were not familiar with as a baseline.

by prior studies [182]. The results of this survey are illustrated in Figure 5.10 B). Additionally, we found that inexperienced group (i.e., raters with experience less than or equal to 4) skews more in favor of assigning unsafe ratings to randomly surveyed locations. This follows our expectation that less experienced cyclists are less likely to feel safe and are more likely to benefit from a map of cyclist safety. This is shown in Figure 5.11. As a comparison, Figure 5.11 also shows the distributions of safety ratings assigned by the experience group for locations they were familiar or unfamiliar with. The increased overlap in the safety ratings assigned between familiarity groups indicates that familiarity with the specific location was less of a factor in influencing the participants assigned safety rating than their own experience cycling.[4]

After collecting survey responses, we compared the perception of safety at each location to the presence of close passes within a distance margin. To do this, we considered locations with an average perceived safety rating above a 4 as "high risk" and any locations less than or equal to 4 as "safe". Because of the limited agreement across survey responses and limited quantity

[4]The experience participants were familiar with roughly half (783 of 1617) locations rated by them.

of survey responses in general, we were unable to generate a correlation with significance of $P<0.05$. Despite this, when comparing the locations with close passes to the locations with perceived risk, we find a PR-AUC of 0.85. The results at other distance margins are available in Table 5.1. This indicates that locations where close passes were observed are more likely to be perceived as "high risk" on average while locations where no close passes were observed are more likely to be perceived as "safe".

Perceived Safety as an Indicator of Collisions

For completeness, we correlated our perceived safety survey responses to collisions as a baseline. This results in a PR-AUC of 0.82, indicating that close passes are similarly accurate to perceived safety surveys, but are significantly more scalable with little to no user effort and can provide continuous views into road safety as infrastructure changes. For example once ProxiCycle is installed, close passes can be sourced entirely passively without user input and continuously over time while perceived safety requires routine user reporting which does not scale. For context, our perceived safety surveys required approximately 30 minutes to complete for 50 locations, which across 76 users is almost 40 hours of labeling, while ProxiCycle requires only minutes to install. Additionally we found close passes measured using ProxiCycle to to have a higher correlation with higher significance to 5-year collision data than perceived safety survey responses.

5.5.2 Participant Feedback

At the end of the deployment we ask all participants to join an optional 20-30 minute exit interview. 8 of the 15 participants could attend and accepted the interview. These 8 participants accounted for nearly 80% of distance traveled during our study. Prior to the interview study we shared a custom interactive map visualizing each user's routes and associated close passes to reflect on prior to the interview. The interviews were semi-structured and open-ended, conducted remotely over Zoom while the user's map was screen shared. Each participant was asked to share their experience using the sensor as well as their thoughts on the resulting close pass hot spots (specific to their own routes) accumulated over the study. Participants were also asked whether they thought the hotpots indicated anything about their

	Estimating Collisions				Estimating Perceived Safety				Perceived Safety Baseline			
Distance Resolution	**Corr**	**ROC AUC**	**PR AUC**	**F1**	**Corr**	**ROC AUC**	**PR AUC**	**F1**	**Corr**	**ROC AUC**	**PR AUC**	**F1**
100m	0.40	0.71	0.63	0.69	–	0.52	0.74	0.49	0.12	0.58	0.62	0.57
150m	0.51	0.74	0.74	0.75	–	0.52	0.82	0.46	0.15	0.60	0.76	0.56
200m	**0.60**	**0.79**	**0.83**	**0.80**	–	0.53	0.85	0.46	–	0.60	0.82	0.53
250m	0.71	0.85	0.87	0.84	–	0.52	0.88	0.44	0.16	0.61	0.88	0.51

Table 5.1: A table showing the performance of using locations of close passes to estimate both locations of collisions and perceived safety as measures of risk. The last column presents baseline measures of using perceived safety survey responses as the estimator of collision locations. Metrics are shown at different distance margins for spatial aggregation to demonstrate performance at different resolutions. Pearson's correlation is provided for all comparisons. Correlation is omitted in comparisons with P>0.05. F1 score is calculated using Macro-F1.

rides and if so what? We also asked them to identify who might benefit from the insight from this close pass hot spot map and why? The interviews were conducted by a single author who made notes and used Affinity Diagrams to generate themes across participants [183]. We quickly reached saturation [184] and 3 themes emerged from these interviews.

Theme 1: Utility to Novice Cyclists, policy makers, and planners

A common theme across participants was the utility to primarily novice cyclists and potentially policy makers. *"Absolutely, this would be very informative for people who are new to biking in an area and haven't figured out their own route preferences yet, especially in places like Seattle where it's all about trade off's like how much topography you're willing to deal with. In order to find your routes, you need to see objective data and the more data you have the better."* (P6) Another participant recalls their time as a beginner cyclist mapping the city themselves *"I would say if I hadn't ridden these routes before, and I saw a map like this, it would impact where I might choose to ride... For example, when the bike lane got put up on Roy, I literally didn't know about it so I would ride on Mercer which is really not a great place to ride. Programs like Strava where they show you where people ride is nice, but it's not clear if it's just a race or actually people riding"* (P2). This raises the importance of safety context along with empirical measures of route popularity. Some participants noted that close passes may be more indicative of rider comfort than true safety: *"I think this data would be useful because close passes don't necessarily hurt anyone but they contribute to a general feeling of unsafety. There's right hooks and dooring which may be more severe crashes but your data is capturing something that impacts more of how people choose to interact with biking"* (P5). This again highlights the importance of comfort and perceived safety on lowering the barriers to cycling adoption.

Theme 2: Impact of Rider and Driver Behavior on Utility

Most participants raised awareness to potential confounds due to their own behavior during data collection. For example, participants raise concerns of the lane position they ride in leading to closer or further passes. *"I was actually hit by a car on Roosevelt but since then my personal choice has been to take more quiet routes at the expense of time. My bike style is very much ride in the middle of the lane. My personal data points probably aren't the most useful, but with more data and more riders, the utility of the data might improve"* (P4). Another participant raises the opposite consideration with their data *"I am actually more biking off to the side of the lane because that's where I feel safer so that's why there are fewer close passes than expected here."* (P3). Both of these participants identified behaviors that may lead to fewer close passes compared to their expectation of an average cyclist due to their riding style, despite citing opposite factors (i.e., taking up more or less of the lane). Users also noted variation in driver behavior across time leading to different results. *"A lot of bad places, no close passes showed up. But I do feel like cars were giving me a wider berth. I don't know if something has changed at the beginning of the month or a coincidence of the time. But I was hoping the sensor would capture this and it just didn't happen even though it has happened dozens of times over the last few years"* (P1). Similarly participants noted the time of day impacting the signal. P5 and P3 raised the fact that they rode the most dangerous routes in the early morning when few cars pass, leading to the illusion of safety on this route if time of day is not considered. Finally, some participants noted the importance of driver speed to contextualize close pass data. *"Another thing is the faster I'm going, the less I care about cars passing me because they are passing me at a similar speed. New cyclists are likely going slower and so it must be much scarier for them"* (P5). Another participant notes, *"Keep in mind that downtown things generally move slower so close passes are less scary. In suburban roads which are paint only and higher speeds, behavior dynamics are different. I'm more assertive and take the whole lane here but close passes here are much worse due"* (P6). This feedback demonstrates the importance of looping in temporal and environmental context whenever using close passes for downstream analyses.

Theme 3: Ease of Data Acquisition Making Up For Errors

Despite multiple participants raising concerns about potentially misleading information, all participants agreed that the data was valuable and easy enough to collect to merit using the system. *"There's definitely some improvements but it's easy enough to do and I'm fairly motivated to collect data because I see It as useful"* (P5). Another notes, *"I'd absolutely keep using this. I find it pretty easy to use and I kind of treat it like my lights. I already charge things and take things off my bike when I'm not riding. It kind of fits into my routine"* (P4). Interestingly, the use of multiple bicycle mounted tools is a trait common amongst experienced cyclists, further motivating this tool as something experienced cyclists can use to record data which can be shared with novice cyclists to propagate expertise amongst the community.

5.6 Discussion

In this section we will discuss ideal applications for ProxiCycle along with an action plan for scalable adoption. Then we discuss the limitations of first sensing close passes and then utilizing close passes for inferring safety across the road network, along with opportunities for future work.

5.6.1 Application Scenarios

This work demonstrates the feasibility of deploying ProxiCycle to study cycling safety in the absence of historic safety priors and on a shorter timescale than existing approaches, enabling quicker and more continuous evaluation.

Routing for Novice Riders and Outreach

We envision a number of applications which this affords. First and foremost, ProxiCycle can be used to bootstrap bicycle safety studies in municipalities lacking any records of reported bicycle collisions, perceived safety, or rides. Cycling activist groups could deploy ProxiCycle with a few riders to estimate safety across the road network which can be used to advocate for infrastructure improvements. This signal can then be used in navigation applications to direct cyclists on known safer routes or can be provided as a map layer for data exploration. These mapping tools could then be used for education and outreach to promote cycling as related work [145] and our formative study in Section 5.2 indicate that knowledge of safer routes can be the deciding factor for some people on whether to take the first step into cycling. ProxiCycle can also provide useful feedback to novice riders on the safety of their chosen routes to expedite the development of their mental maps.

Planning

Additionally, we envision ProxiCycle being useful for A/B testing navigation routes as systems built alongside ProxiCycle could suggest alternative routes and automatically measure the safety across them using the quantity of observed close passes on each route in real time. Similarly, data gathered by ProxiCycle users before and after infrastructure development can be used in empirical statistical analysis to quantify the impact of bicycle related development on bicycle safety. Empirical data like this can help support future investment into healthy infrastructure by demonstrating its immediate impacts on safety.

Collision Modeling

Further, since close passes occur much more frequently than collisions, they can be used in data-hungry analyses which may be infeasible with more scarce collision data such as comparing safety across time of day. Similarly, the locations of close passes could be used to train data-hungry models to estimate problematic locations or even predict future risk at scale. These models could then be used for more dynamic safety-based navigation which accounts for changes in bicycling safety over time on shorter timescales that what is possible with collision data.

5.6.2 Adoption Plan

Like any crowdsourcing system, ProxiCycle depends on user adoption. In Section 5.5 we saw the potential for grassroots neighborhood organization to bootstrap adoption. Similarly, much of the original location-aware computing work leaned heavily on hobbyist "War-Drivers" to map wireless networks. [185]. In Section 5.5.2 we found our participants valued contribution to sourcing close pass data worth the effort to ride with ProxiCycle . While ProxiCycle devices could be built by hobbyist or distributed by activist organizations, we also see room for engagement within the private sector through distribution from insurance companies or bike-share companies. Many organizations from governments to private companies (often in collaboration with governments) offer economic incentives for bicycle transportation from discounts at local businesses all the way up to e-bicycle purchase rebates[5]. A similar program could exist for distributing ProxiCycle stakeholders with an interest in understanding safety (i.e., governments, insurance companies, etc.) offering rebates in return for volunteering to use ProxiCycle on rides.

5.6.3 Limitations & Future Work

While this work demonstrates the feasibility of passively crowdsourcing close passes as a measure of road safety or comfort, there are multiple limitations and opportunities for future work. Primarily, close passes offer a view into the risk of a certain type of collision, so-called "side-swipes", but may not represent near-collisions of different formats such as "dooring" or "right-hook" collisions. Analyses utilizing close pass data may therefore benefit from additional safety context. Similarly, historic traffic levels were unavailable at the time of this analysis, but future studies could leverage traffic data to studying the weighted correlation in close pass and collision rates with respect to total trips at each location. Additionally, since the absence of a close pass does not inherently mean 0% risk, this analysis can characterize the stability of close pass rates over different time windows.

The performance of close passes as a measure of safety may also differ across road type based on bicycle infrastructure treatment, road width, and other features. A follow-up analysis could investigate the correlation of close passes with existing safety measures across road type and the presence of cycle infrastructure or other factors such as time of day or weather conditions. Similarly, additional sensors could be introduced such as low-resolution cameras or micro-

[5]Anecdotally, one of our participants noted during our interview that their employer offers a program which funds a bicycle purchase and routine maintenance for the employee if the employee commits to commuting to work at a minimum of twice a week

phones to measure qualitative differences in the types of vehicles committing close passes such as size or speed to further investigate the severity of close passes at different locations, i.e., larger and faster vehicles are strictly more dangerous [186]. Finally, future studies could incorporate cyclist activity recognition to incorporate cyclist behavior context (i.e., lane position or riding speed) which may alter the performance of close passes as an effective collision surrogate.

5.7 Conclusion

In this chapter we address **RQ.3** which asks: how can the paradigm of early-detection in public health sensing extend beyond symptoms and biomarkers to detect society level behaviors which pose safety risks, and can these insights be used to mitigate these risks by providing aggregate safety feedback?

We address this through the design and development ProxiCycle , a standardized and scalable approach for crowdsourcing a measure of cyclist safety along the road network by sensing the proximity of passing cars, in service of developing and evaluation a novel collision surrogate to passively measure cyclist safety before anyone gets hurt. We motivate the development of this system through a formative study to confirm the need for a cycling safety signal. We then validate the system in both an in-lab and real-world environment to demonstrate its accuracy at both close pass detection and legal pass detection. We finally deploy the system for 2 months with 15 cyclists over 1,000 miles travel to accumulate a passively crowdsourced map of close-passes, enabling the evaluation of close-passes as a collision surrogate. We then compare the close passes recorded by this system during real-world use against both historic bicycle collision data and user-reported perceived safety finding a significant Pearson's correlation of 0.6 in locations of close passes in just 2 month of deployment with 5 years of collision data. We show that these close-passes can estimate empirical cycling risk across the road network in the absence of existing historic collision, can converge in less time, and serve as a predictive measure of safety which does not depend on collisions occurring. This work mirrors the early-detection paradigm of Chapter 4 but for a different side of public health - injury prevention. This work leverages the ubiquity of smartphones already deployed across the road environment as a platform to extend with minimal hardware for sourcing a high impact signal for public health. We then show through our qualitative evaluation that participants derive utility from the resulting close-pass map for safety based routing and even postulate that such data could be useful for planners and policy makers. We believe this work is a valuable step in advancing a line of inquiry into bicycle safety by demonstrating the feasibility of ProxiCycle as a fully passive and easily scalable sensing platform for measuring cycling safety across the road network.

Chapter 6

Opportunities of Ubiquitous Sensing Platforms for Common Good

Taking a step back, I would like to discuss some of the lessons learned through these works and highlight further opportunities to benefit the common good with ubiquitous sensing. I will discuss these ideas within the context of the work in this thesis.

6.1 Advantages of Ubiquitous Platforms

Surrogate signals are by definition a stand-in for something else. Ideally, we would sense that *something else* directly, as it is the most direct signal of what we'd like to measure. However, we use surrogates because this *something else* has obvious limitations. In most cases, the limitation has to do with the difficulty of accessing the signal, limiting scalability, or the delay between the events and when the measured signal is available, limiting utility. For example, diagnosed fever might require purpose-built hardware to access, limiting scalability, while road injuries simply do not happen often enough to make scaled analyses possible. We can address many of these limitations with sensing surrogate signals and many of the advantages of ubiquitous platforms – like scalability, accessibility, longitudinality, and physicality – complement the needs of a good surrogate signal. I will now discuss each of these strengths of ubiquitous computing platforms in the context of surrogate sensing below.

6.1.1 Scale: From the Individual to the Population

The most obvious benefit of developing surrogates on top of ubiquitous sensing platforms is that most of, if not all, the infrastructure to leverage them is already there. When we use side-channel sensing, we trade in a small decrease in inference accuracy for a significant increase in the scale of impact. This is particularly important in public health challenges where data coverage can have an even stronger impact than individual measurement accuracy. This is because applications in public health often benefit from scale surpassing a tipping point, similar to herd immunity in the context of vaccines, and sensing is no exception. Where personal sensing might capture actionable individual insights given sufficient accuracy, slightly less accurate population scale screening can afford change at the institutional level given sufficient scale is met (i.e., study of societal behavior or uncovering environmental factors which can be addressed using population scale data). Additionally, change at the institutional level is extremely powerful as the immediate impact is shared across the population, but also sets the context for continued positive change like a feedback loop. I discussed this in the context of promoting cyclist safety in Chapter 5 where population scale insights about the road environment can enable changes to policy or physical changes to the road environment that improve the safety where it is needed most. Not only can this impact the riders directly by improving the safety of those spaces, but also sets the stage to encourage more people to adopt cycling, which further improves cycling safety through safety in numbers. Similarly, while accurate flu tests might be used to confirm a test, passive and continuous immune response detection across a population could enable pre-screening that could address limitations of flu surveillance due to sample bias where infected individuals never know to test or improve timely geospatial resource allocation during epidemics.

More abstractly, and at a much greater scale, there is a possibility for ephemeralization[1] of population sensing where scaled data collection from ubiquitous computing leads to empirical insights that fuel studies that eventually result in common knowledge or best practices. At this point population sensing itself becomes obsolete and all that's left is to simply follow a set of guidelines which are understood to solve the problem that the sensor was designed to assist in, should they be followed strictly (e.g., discovering good urban design patterns which

[1]A term coined by R. Buckminster Fuller in 1938 to describe how advanced technology continually achieves "more and more with less and less until eventually it can do everything with nothing".

can be implemented herein, discovering healthy lifestyle choices that work for specific subpopulations, etc.). This trajectory of ephemeralization was formalized in population health as the "Translational Spectrum" [187] in which a solution to a health problem translates first from a research bench to a person (T1), then from a person to a real patient (T2), then from a real patient to being used in practice (T3), and finally from practical local settings to a population context more broadly (T4). The work described in this document aims to align ubiquitous computing platforms with the translational stage, T4, and enable population scale insights to establish, improve our understanding of, or support the recognition of social determinants of health. That is, in some cases, the end goal is not to measure these signals, but to leverage signal measurement as a tool in developing a holistic understanding of a problem in-situ. Though in practice, having a continually up-to-date empirical dataset that captures changes in the population over time can have additional benefits as environments and populations are not static.

6.1.2 Delay: Real-Time Epidemiology

The other benefit of these systems is reduced delay in response. Traditional systems for sensing meaningful signals in epidemiology are subject to a number of delays including physiological (i.e., delay in an individual's perception of their own symptoms), behavioral (i.e., delay in an individual's decision to seek testing and care), and systemic (i.e., round-trip time for a test to be taken, processed, and resulted received) delays. Passive and longitudinal sensing may address each by removing the need for human perception and decision, thus eliminating physiological and behavioral delays, and performing this sensing on connected digital devices which can recover and share data instantly, thus eliminating systemic delays as well. I discussed these ideas in both Chapter 4 and 5. Specifically chapter 4 builds on this by addressing physiological delays by passively measuring signals in the body that could indicate immune response to infectious disease even before an individual becomes aware they are getting ill. Systems like this offer an opportunity to close the loop between epidemiological study of population scale datasets and the sensing systems that source them. Specifically, ubiquitous platforms for sensing afford the ability to run epidemiological models in near real-time.

Real-Time Epidemiology In The Context of Infectious Disease

I argue that Chapter 4 is valuable work towards a future of real-time epidemiology for ILI. At the personal level, smartwatch users can receive updates informing them that they are showing signs of infection and pursue a PCR test or even an affirming fever check with their smartphone. However, at the population scale, the information that the flu is likely going around could be shared throughout media outlets days, or even weeks, before traditional means would have uncovered it, allowing other people to take preventative action early to minimize exposure. Closing the loop, the same population scale model of disease likelihood could serve as a Bayesian prior for the same passive biomarker models that sourced them (i.e., integrating population state information to the individual models).

This is in contrast to existing contact tracing systems which, despite their use of ubiquitous platforms, still depend on slow PCR tests. Because these systems do not leverage passive sensing as part of the illness detection, they are bottle-necked by the same behavioral and systemic delays as traditional systems.

Real-Time Epidemiology In The Context of Injury Prevention

I maintain a similar argument for Chapter 5 in that a rolling buffer of measured close-passes across the city can provide a constantly updating view of cycling safety. Urban environments are far from static. They develop or fall into disrepair. Roadside construction can close down cycle lanes or detour arterial road traffic onto residential side streets changing the cycling safety landscape overnight. While current approaches of geospatial analysis may capture some variation over time, they are limited to what is represented historically via infrequent snapshots. It is possible to model and predict the effects of changes, but these models are based on assumptions which may not hold as other artifacts of the environments change as well. Models based on continuous longitudinal signals like close-passes can be repeatedly validated in new environments, enabling a more up-to-date understanding of the evolving environment, creating something like a road safety time series for the entire road network.

You can imagine the following scenario in which ProxiCycle is deployed in a city with limited cycling infrastructure. Over time the statistics on close passes at different points along the road network emerge from the aggregated historic data. This data is regularly updated on an open mapping platform which can integrate with existing navigation tools to provide routes which are optimized for safety, among other features, based on data from ProxiCycle. Inexperienced cyclists who are unwilling to make the blind leap to cycling on the road can check this platform for safer routes both with and without existing bike infrastructure. These inexperienced cyclists are able to identify convenient routes for at least some of their trips and feel secure to adopt cycling as a safer transportation option. The growth in the cycling community demonstrates a need for improved infrastructure which planners and policy makers can prioritize based on the same data generated by ProxiCycle. As infrastructure is introduced, cyclists begin using it and the data generated by ProxiCycle quickly converges to reflect this new reality, both informing inexperienced cyclists about this new route and giving instant feedback to planners and politicians on the impact of their intervention. The data from this naturalistic study before and after intervention can then be used to vouch for future interventions to continue improving the safety of the city.

6.1.3 The Physicality of Sensing

A final benefit of ubiquitous sensing over other approaches is physicality. While one could argue that a correlation between road design and historic collisions is not causal, it is hard to argue that the proximity of passing cars is not inherently linked to the likelihood of side-swipe collisions. A collision is ultimately just a close-pass at the closest proximity. This is the basis for Heinrich's Triangle of injury prevention – a well established theory that supports the use of near-misses as safety surrogates [162]. Sensing offers a grounding in physical intuition that other inference based methods lack. While the aforementioned benefits of scale and reduced delay are results of ubiquity and connected digital computing, this benefit of physical grounding is a result of sensing as the approach of sourcing our independent variables. Sensing a signal that quantifies an outcome is a first-principles approach to modeling and I find this to be an increasingly overlooked quality in the age of data mining and learning unexpected higher-order relationships with AI. While some of these projects like Chapter 3 leverage AI to learn mappings from one signal to another, they all still build off a physically grounded sensor modality that is the exact tool for the application. Machine learning simply works like a filter to boost signal-to-noise ratio.

6.2 Extending Ubiquitous Computing Platforms for Public Health

Aligning sensing on ubiquitous computing platforms for sensing in public health is often indirect and requires additional creativity. The 3 studies that compose this thesis leverage 3 unique approaches of extending ubiquitous computing platforms to achieve this alignment. In Chapter 3 I use the constraint of an active interaction to sufficiently narrow a sensor signal so that a reasonable mapping can be learned to achieve side-channel sensing of a useful health metric. Additionally, I introduce additional sensors as supporting features to further improve this mapping. Chapter 4 takes a different approach by using the sensors in a smartwatch fitness tracker as intended – to measure biomarkers – but as surrogates for immune response. In this study, I extend these sensing platforms by specifically studying their response at the earliest stages of infection. This work builds on the strength of wearable devices being always available as an opportunity for early-detection of ILI. This work does not change *what* is being measured by the sensor, but *when* it can be. The unique approach leveraged in this study is the introduction of a new ground truth in the time since exposure. The sensor hardware and even software processing remains the same, but a novel frame of reference is introduced to assign new value to the same device. Finally, in Chapter 5, I extend smartphones by using them as the computing backend for a small set-and-forget sensing device. While this approach does require additional hardware, it makes use of an existing human activity to passively sample across space and the existing compute in the rider's pocket to communicate that data in real-time automatically. I encourage future researchers and developers to lean into these approaches:

1. the use of physical interactions and additional context signals to enable new applications of existing sensors (i.e., side-channel sensing),
2. careful consideration of the frame of reference in relation to the decision boundary (i.e., moment in time relative to a measured event, granularity, etc.), and
3. opportunities to piggyback sensing on top of existing human activities to enable convenient passive sensing, even at the expense of additional hardware.

6.2.1 The Ground Truth Bottleneck

Leveraging these approaches typically comes in the form of introducing new ground truth to contextualize a signal. In Chapter 3 the sensor was not modified, but rather a very particular ground truth was introduced: the core body temperature of the user during an active physical interaction. This highly curated ground truth enabled learning the mapping that disaggregates the core body temperature from the sensor readings. In Chapter 4, the sensors were again used as intended, but within the context of a prospective study environment in which simply knowing the moment the study participants were exposed to the flu contextualizes all the samples in their continuously measured data (alongside even more thorough clinical tests). Finally in Chapter 5, the added ground truth comes in the form of historic collision data by running the data collection within municipalities which have well documented injury statistics across space. In none of these studies were the fundamental sensing methodologies the bottleneck. It was rather the complex study design that enabled re-contextualizing the same sensor signal within the resulting ground truth that developed on the sensing paradigm.

6.3 Next Steps

In this section I will cover some opportunities for future study and my perspective on how to pursue population scale sensing with ubiquitous systems in the real world.

6.3.1 Support for Manufacturers and Corporations

Studies like those presented in this thesis serve as first-order explorations of what can be possible given further support from manufacturers. I posit that demonstrating compelling applications like FeverPhone should get device manufacturers excited about exposing more sensor data to developers. In fact, shortly after publication, Google Pixel shipped an external IR temperature sensor for approximate fever detection, illustrating manufacturer appetite for these use cases once demonstrated. I make a similar argument in Chapter 5 for the adoption of ProxiCycle by ebike-share companies. Micromobility companies like CitiBike [188] and Lime [189] have created the ideal platform for sensing geospatial factors that could affect the active transport experience in-situ. By rolling out new - or re-contextualizing existing - hardware, these companies could quickly create living maps of the road network that can reflect the current environment even with frequent modifications.

There are countless other population scale signals to be explored in this way that can provide public utility. As one example, following the completion of ProxiCycle, I led a similar study demonstrating that the ambient light sensors in smartphones held by pedestrians while walking can effectively crowdsource and map nighttime ambient lighting conditions to improve nighttime pedestrian route planning in navigation apps [190]. This work was motivated by the stark lack of data representing lighting conditions on navigation platforms despite the importance many people assign to it as a deciding factor in *where* or even *whether* to walk at night.

This example is particularly interesting as it provides no real utility to the user collecting the data until it is aggregated across many other users at population scale. Unlike ProxiCycle, which could at least give users a better sense of how close cars are actually passing them on their routes, this work – we call NightLight – simply tracks the brightness measured by the user's smartphone's ambient light sensor. Where ProxiCycle quantifies something that is hard for humans to really perceive (especially while their attention is occupied by cycling), NightLight quantifies lighting conditions, which are much more intuitive to understand even without quantifiable data. NightLight is also purely software defined, making it an extreme example of how a small amount of software (i.e., recording sensed light) can facilitate a small change to an existing platform (i.e., mapping sensed light geospatially in navigation apps) which could have a small impact on individuals (i.e., making walking at night more accessible) but can result in significant impacts when measured across the population (i.e., increase physical activity and therefore cardiovascular health across the population [32]). These small changes even have even further reaching impacts in that they quite possibly also initiating a feedback loop in that increasing foot traffic also increases "eyes on the street"[2] which in turn again increases foot traffic [191]. While creating an entirely new navigation platform specifically geared towards pedestrians who are concerned with nighttime lighting conditions might be too much work for the returns, these software changes to existing platforms are minimal.

[2] A concept coined by Jane Jacobs in her seminal work on Urban Planning *The Death and Life of Great American Cities*

Corporations and existing platforms should therefore explore applications of opportunistic sensing which can leverage the approaches outlined in Section 6.2 to provide public utility as they are oftentimes the gatekeepers of scale.

6.3.2 Demonstrating the Effects of Scale

Another exciting opportunity for research is the investigation of the nonlinear effects of scale in these ubiquitous or collectivist systems. For example, a combination of signals at the personal level may not provide any additional insight, but what if the same combination of signals, when sampled across a population, reveals an entirely new compositional effect? Returning to the epidemiology example, early-detection of influenza cases may lead to personal behavior change to protect others, but at the population level, a geospatial signal of early-immune system response could uncover environmental factors like superspreader events even before the first PCR test and facilitate immediate institutional response. Uncovering and demonstrating these population scale effects requires deployment and longitudinal studies similar to Chapter 5 across new domains. For example, what can geospatial trends in relative heart rate or heart rate variability indicate about a population? Can the spread of catastrophic news throughout the country be traced with population scale cardiovascular effects measured by wearables? On a more local scale, can problematic street intersections that result in frequent near-misses between cars, cyclists, and pedestrians be identified by transient spikes in heart rate by smartwatch users in these exact coordinates?

6.3.3 Alignment with Existing Institutions and Incentives

Translating these ideas from academic studies to real world systems hinges on collective adoption and central planning. This is most directly achieved through support from a governing body (i.e., policy makers), top-down economic incentives (i.e., insurance companies), or existing non-profits or activist groups. Take for example the success of Project RISE - a collective effort to sense and expose environmental injustice committed by local materials processing plants [192]. This work was made possible through organized effort of an environmental non-profit. Leveraging technology to quantify something is one thing, but extracting real value from it requires building complex sociotechnical systems around them that align the technology with a political or community incentive. I raise the idea of economic incentive because we already see something like this in the private auto insurance sector where drivers are offered insurance payment discounts in return for evidence of good driving behavior through the lens of opt-in sensor data collection [193]. You can imagine a system like ProxiCycle providing a similar opportunity[3].

Within the context of using commodity smartphones and smartwatches for health diagnostics, these solutions hinge on a collective buy-in from manufacturers, but more importantly regulators, to create the standardized conditions for these systems to generalize cross-device. In Edward Jay Wang's thesis conclusion he notes that *"The ubiquity of this solution space hinges on shifting the central definition of the 'smartphone' to include health and medical sensing as a use case no less important than making a call or connecting to the internet"* [194]. I propose something similar, with a focus on standardization to optimize population scale aggregation. Organizations like the Food and Drug Administration (FDA) which regulate and approve health devices could propose standard regulations that smartphone manufacturers could follow to gain expedited approval for medical use-cases. Not only would this incentivize a more standardized set of sensors, making cross-device generalization easier, but would also establish a market for medically focused lines of smartphones. Imagine a version of a smartphone with specialized components to enable diagnostic tests that could be sold to healthcare providers and community health workers.

6.4 Closed Loop Sensing

While this thesis introduces concepts of real-time epidemiology, none of the projects described achieve this directly. While Chapter 5 comes close through a longitudinal deployment, it lacks a demonstration of naturalistically validating safety effects of updated road designs on the population. The most important next step is to move towards fully online inference in which interventions are administered due to signals captured through population scale sensing and the effects evaluated in the same environment across time.

[3] There is obviously a lot of complexity to this which is out of scope of this dissertation but I caution that economic incentives tend to fall victim to Goodhart's Law which states "Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes." This isn't to say it's not possible, but that it is more complex than I can discuss here.

6.4.1 The Sociotechnical AI flywheel

In Section 6.2.1 I note the dependency on novel ground truth signals to align these sensing systems with new use cases. I argue that one of the most substantive next steps in this space is the automation of ground truth collection through AI. Language Models (LMs) increasingly replicate human-expert performance on language reasoning tasks. For example, in SymptomAI, LM agents were shown to outperform experienced clinicians at differential diagnosis given symptom descriptions communicated by laypeople. After validating their accuracy on a small subsample, these AI generated differentials were used to conduct a phenome-wide association study (PheWAS) of over 13,000 users' wearable data, mapping biosignal trends to AI-diagnosed conditions [195]. These types of auto-scaled analyses, leveraging AI-generated ground truth, may be increasingly common as LMs lower the barrier of entry to sourcing population-scale label sets. Though, these labels must still be grounded in some context to ensure they remain strictly a transformation of an independently collected feature set (i.e., symptom descriptions mapped to diagnoses) and not a transformation of the independent variables themselves (i.e., deriving diagnosis from wearable data directly with an LM).

Sourcing this type of pseudo-ground truth through AI-assistance could quickly become a flywheel for learning these types of signal mappings by allowing the very same study participants who provide the sensor data to also provide the label through accessible language reporting (i.e., coupling wearable data with daily health and activity diary). Particularly if the users themselves begin to recognize patterns in their wearable data through the expertise provided by LMs processing their journals. This may present yet another virtuous cycle in which the sensor data helps inform the LMs understanding of the user and the LM's general knowledge can educate the user in ways that may not have been possible alone. Over time, the combination of these experts – the individual and the generalist LM annotator – develop a shared expertise that improves the quality of ground truth labeling of sensor data herein. This type of flywheel adoption depends on sociotechnical alignment of AI in ways that support independent human creativity, reasoning, perception, and education in order to propagate domain expertise from generalist LMs to laypeople and achieve civilian labeling at scale.

6.5 Computational Collectivism

Ending on a philosophical note, a common theme among the projects in this thesis is leveraging ubiquitous computing platforms in applications that benefit everyone, not just the user themselves. Chapter 3 empowers individuals to assess their own health in the context of infectious disease which can benefit those around them through allowing them to effectively mitigate spread. Chapter 4 follows this same theme of disease mitigation, but from a prognostic standpoint leveraging predictive leading indicators sensed passively in the background. This time users do nothing more than share their data and receive a potentially healthier community in return. Chapter 5 similarly senses a predictive measure, but in the domain of injury prevention. In this case, a small group of individuals crowdsource and map a useful geospatial signal that predicts road safety for all other cyclists on the road. Not only does this directly improve the safety of those who ride those routes, but it additionally can lower the primary barrier of entry to cycling which allows a greater population to reap the health benefits cycling has to offer. Both these applications allow users to take on minimal individual effort for the good of the community.

Sensing often gets a bad reputation in the public eye because of its potential for surveillance. This is not unfounded as systematic inequality has been uncovered in data-driven systems like personalized advertisement [196]. But we've also seen that people are fairly quick to opt in to surveillance programs when they provide a modest amount of individual convenience, like not having to stop at toll booths on the highway [197] or allow you to tap a credit card instead of carrying around cash.

Projects like ProxiCycle [198] and NightLight [190] are a bit different in that the majority of the benefits they provide are to others. A user does not get any immediate benefit for tracking close-passes or ambient light levels, but do have to share sensitive location data. If they were strictly concerned with personal benefit they might opt for a collision avoidance system or a flashlight. But projects like these allow users, with very little effort, to contribute to a collective effort to improve navigation platforms for all. These collectivist systems do eventually benefit the user with time though, as mapping these values can lower the barrier of entry to things like walking or cycling, which should grow the populations that leverage them and create the conditions for further improvements (i.e., paving new cycle lanes and installing infrastructure to facilitate comfortable nighttime walking). There are a number of barriers limiting the feasibility of these types of collectivist applications, namely political acceptance and overcoming individualistic ideology. However, I anticipate once society is ready, they will have a substantial impact at scale.

Chapter 7

Conclusion

This thesis set out to demonstrate that sensors on ubiquitous computing platforms like those already built into smartphones and wearables can be extended or repurposed to improve public health and well-being through early detection, or even prediction, of illness and injury at the population scale. I approached this claim not by building new sensing hardware, but by finding new value in hardware that society has already deployed by the billions.

The three studies that compose this document each move a step further along the timeline of public health intervention. In Chapter 3, FeverPhone re-contextualizes a smartphone's internal thermistors through a brief physical interaction to make fever screening – a diagnostic – potentially accessible to anyone with a phone. In Chapter 4, a first-of-its-kind challenge study shows that the biomarkers already measured by commodity smartwatches respond to influenza in its earliest days, even for mild, asymptomatic, and PCR-negative cases, pushing detection into the pre-symptomatic range. And in Chapter 5, ProxiCycle extends the smartphone into a passive backend for crowdsourcing close passes as a collision surrogate, mapping cyclist safety before anyone is harmed. Together they trace a path from diagnosis, to prognosis, to prevention.

Reflecting across these works, the most durable lesson was not about any single sensor. In each case the sensing itself was rarely the bottleneck; the difficulty, and the innovation, lay in the study design that introduced the right ground truth to give an existing signal new meaning – a curated interaction, a known moment of exposure, a municipality's collision record.

This suggests that as ubiquitous platforms continue to lower the cost of *access* to signals, the frontier of this line of work shifts toward how we *ground* them, and how we close the loop from measurement to intervention in the environments where these signals are collected.

Finally, the three projects share a quality I have come to see as their most important. Each asks the individual for very little – a ninety-second measurement, a shared data stream, a sensor left plugged into a handlebar – and returns a benefit that accrues largely to others and to the community as a whole. This is the promise of what I have called computational collectivism: that infrastructure already in our pockets and on our wrists can be aligned, with modest effort, toward outcomes no individual could achieve alone. Realizing that promise at scale will depend as much on institutions, incentives, and public trust as on the sensing itself. But I believe the work presented here is a meaningful step toward a future in which the ubiquitous computing already woven through everyday life quietly helps keep people healthier and safer than they could be on their own.

Bibliography

[1] Pew Research Center. Mobile fact sheet, November 20 2025.

[2] Jeffrey Gottfried, William Bishop, Monica Anderson, Michelle Faverio, Eugenie Park, and Colleen McClain. Americans and ai 2026: Chatbots, smart devices and views on impact, June 17 2026.

[3] Eric Lansdown Trist and Kenneth W Bamforth. Some social and psychological consequences of the longwall method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Human relations*, 4(1):3–38, 1951.

[4] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *science*, 368(6491):eabb6936, 2020.

[5] United Nations. The global goals. 2015. `https://www.globalgoals.org/`.

[6] Jan Gehl and Birgitte Svarre. *How to study public life*, volume 2. Springer, 2013.

[7] CDC Foundation. What is public health?, 2025. Accessed May 26, 2025.

[8] C-EA Winslow. The untilled fields of public health. *Science*, 51(1306):23–33, 1920.

[9] Max Roser. Causes of death globally: what do people die from? *Our World in Data*, 2021. https://ourworldindata.org/causes-of-death-treemap.

[10] CDC. Estimated influenza-related illnesses, medical visits, hospitalizations, and deaths in the united states — 2019–2020 influenza season. `https://www.cdc.gov/flu/about/burden/2019-2020.html`, November 2023. Accessed: 2024-03-01.

[11] David Blumenthal, Elizabeth J Fowler, Melinda Abrams, and Sara R Collins. Covid-19—implications for the health care system, 2020.

[12] Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.

[13] Kassiani Mellou, Spyros Sapounas, Ioannis Panagoulias, Maria Gkova, Kalliopi Papadima, Anastasia Andreopoulou, Danai Kalotychou, Michalis Chatzopoulos, Kassiani Gkolfinopoulou, Vassiliki Papaevangelou, et al. Time lag between covid-19 diagnosis and symptoms onset for different population groups: Evidence that self-testing in schools was associated with timely diagnosis among children. *Life*, 12(9):1305, 2022.

[14] Isaac Núñez, Pablo F Belaunzarán-Zamudio, and Yanink Caro-Vega. Result turnaround time of rt-pcr for sars-cov-2 is the main cause of covid-19 diagnostic delay: A country-wide observational study of mexico and colombia. *Revista de investigación clínica*, 74(2):71–80, 2022.

[15] Mike A Merrill, Esteban Safranchik, Arinbjörn Kolbeinsson, Piyusha Gade, Ernesto Ramirez, Ludwig Schmidt, Luca Foshchini, and Tim Althoff. Homekit2020: A benchmark for time series classification on a large mobile sensing dataset with laboratory tested ground truth of influenza infections. In *Conference on Health, Inference, and Learning*, pages 207–228. PMLR, 2023.

[16] David A Larsen and Krista R Wigginton. Tracking covid-19 with wastewater. *Nature Biotechnology*, 38(10):1151–1153, 2020.

[17] Hamada S Badr, Hongru Du, Maximilian Marshall, Ensheng Dong, Marietta M Squire, and Lauren M Gardner. Association between mobility patterns and covid-19 transmission in the usa: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(11):1247–1254, 2020.

[18] Andrew Bo Liu, Daniel Lee, Amogh Prabhav Jalihal, William P Hanage, and Michael Springer. Quantitatively assessing early detection strategies for mitigating covid-19 and future pandemics. *medRxiv*, pages 2023–06, 2023.

[19] Liane S Canas, Carole H Sudre, Joan Capdevila Pujol, Lorenzo Polidori, Benjamin Murray, Erika Molteni, Mark S Graham, Kerstin Klaser, Michela Antonelli, Sarah Berry, et al. Early detection of covid-19 in the uk using self-reported symptoms: a large-scale, prospective, epidemiological surveillance study. *The lancet digital health*, 3(9):e587–e598, 2021.

[20] Katayoun Farrahi, Remi Emonet, and Manuel Cebrian. Epidemic contact tracing via communication traces. *PloS one*, 9(5):e95133, 2014.

[21] Mark S Smolinski, Adam W Crawley, Kristin Baltrusaitis, Rumi Chunara, Jennifer M Olsen, Oktawia Wójcik, Mauricio Santillana, Andre Nguyen, and John S Brownstein. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. *American journal of public health*, 105(10):2124–2130, 2015.

[22] CDC. Know the signs and symptoms of infection. `https://www.cdc.gov/cancer/preventinfections/symptoms.htm`, July 2023. Accessed: 2024-03-01.

[23] M Pamela Griffin, Douglas E Lake, and J Randall Moorman. Heart rate characteristics and laboratory tests in neonatal sepsis. *Pediatrics*, 115(4):937–941, 2005.

[24] A Günther, I Salzmann, S Nowack, M Schwab, R Surber, H Hoyer, OW Witte, and D Hoyer. Heart rate variability–a potential early marker of sub-acute post-stroke infections. *Acta neurologica Scandinavica*, 126(3):189–196, 2012.

[25] Aravind Natarajan, Hao-Wei Su, and Conor Heneghan. Assessment of physiological signs associated with covid-19 measured using wearable devices. *NPJ digital medicine*, 3(1):156, 2020.

[26] Denisse Castaneda, Aibhlin Esparza, Mohammad Ghamari, Cinna Soltanpur, and Homer Nazeran. A review on wearable photoplethysmography sensors and their potential future applications in health care. *International journal of biosensors & bioelectronics*, 4(4):195, 2018.

[27] Centers for Disease Control and Prevention. Wisqars. `https://wisqars.cdc.gov/explore/?o=MORT&y1=2021&y2=2021&g=00&t=0&i=0&m=20810&s=0&r=0&me=0&ry=0&yp=65&e=0&a=ALL&a1=0&a2=199&g1=0&g2=199`, 2021. 2024–3-01.

[28] World Health Organization. Sdg target 3.6: Road traffic injuries. `https://www.who.int/data/gho/data/themes/topics/sdg-target-3_6-road-traffic-injuries`, 2024. Accessed: <date of access>.

[29] OECD. Road accidents - oecd data. `https://data.oecd.org/transport/road-accidents.htm`, Accessed 2024. 2024-03-01.

[30] Environmental Protection Agency. Highlights of the automotive trends report.

[31] Monica Dinu, Giuditta Pagliai, Claudio Macchi, and Francesco Sofi. Active commuting and multiple health outcomes: a systematic review and meta-analysis. *Sports medicine*, 49:437–452, 2019.

[32] Tim Althoff, Rok Sosič, Jennifer L Hicks, Abby C King, Scott L Delp, and Jure Leskovec. Large-scale physical activity data reveal worldwide activity inequality. *Nature*, 547(7663):336–339, 2017.

[33] Irene Isaksson-Hellman. A study of bicycle and passenger car collisions based on insurance claims data. In *Annals of Advances in Automotive Medicine/Annual Scientific Conference*, volume 56, page 3, 2012.

[34] Luyang Liu, David Racz, Kara Vaillancourt, Julie Michelman, Matt Barnes, Stefan Mellem, Paul Eastham, Bradley Green, Charles Armstrong, Rishi Bal, et al. Smartphone-based hard-braking event detection at scale for road safety services. *Transportation research part C: emerging technologies*, 146:103949, 2023.

[35] Ahmet-Serdar Karakaya, Thomas Ritter, Felix Biessmann, and David Bermbach. Cyclesense: Detecting near miss incidents in bicycle traffic from mobile motion sensors. *Pervasive and Mobile Computing*, 91:101779, 2023.

[36] Marilyn Johnson, Judith Charlton, Jennifer Oxley, and Stuart Newstead. Naturalistic cycling study: identifying risk factors for on-road commuter cyclists. In *Annals of advances in automotive medicine/annual scientific conference*, volume 54, page 275. Association for the Advancement of Automotive Medicine, 2010.

[37] ST. Vl53l8cx. `https://www.st.com/en/imaging-and-photonics-solutions/vl53l8cx.html`. Accessed: 1-31-2024.

[38] Enrique Hernández-Orallo, Pietro Manzoni, Carlos Tavares Calafate, and Juan-Carlos Cano. Evaluating how smartphone contact tracing technology can reduce the spread of infectious diseases: The case of covid-19. *Ieee Access*, 8:99083–99097, 2020.

[39] André Henriksen, Martin Haugen Mikalsen, Ashenafi Zebene Woldaregay, Miroslav Muzny, Gunnar Hartvigsen, Laila Arnesdatter Hopstock, and Sameline Grimsgaard. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. *Journal of medical Internet research*, 20(3):e110, 2018.

[40] Giulia Regalia, Francesco Onorati, Matteo Lai, Chiara Caborni, and Rosalind W Picard. Multimodal wrist-worn devices for seizure detection and advancing research: focus on the empatica wristbands. *Epilepsy research*, 153:79–82, 2019.

[41] Jennifer M Radin, Nathan E Wineinger, Eric J Topol, and Steven R Steinhubl. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the usa: a population-based study. *The Lancet Digital Health*, 2(2):e85–e93, 2020.

[42] Rosalind W Picard, Szymon Fedor, and Yadid Ayzenberg. Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion review*, 8(1):62–75, 2016.

[43] Maurizio Garbarino, Matteo Lai, Dan Bender, Rosalind W Picard, and Simone Tognetti. Empatica e3—a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, pages 39–42. IEEE, 2014.

[44] Qiuyue Shirley Xue, Yujia Liu, Joseph Breda, Mastafa Springston, Vikram Iyer, and Shwetak Patel. Thermal earring: Low-power wireless earring for longitudinal earlobe temperature sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(4):1–28, 2024.

[45] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3430–3437, 2013.

[46] Xin Liu, Girish Narayanswamy, Akshay Paruchuri, Xiaoyu Zhang, Jiankai Tang, Yuzhe Zhang, Roni Sengupta, Shwetak Patel, Yuntao Wang, and Daniel McDuff. rppg-toolbox: Deep remote ppg toolbox. *Advances in Neural Information Processing Systems*, 36, 2024.

[47] Matt Whitehill, Jake Garrison, and Shwetak Patel. Whosecough: In-the-wild cougher verification using multitask learning. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 896–900. IEEE, 2020.

[48] Qiuyue Shirley Xue, D Shin, Anupam Pathak, Jake Garrison, Jonathan Hsu, Mark Malhotra, and Shwetak Patel. Luckychirp: Opportunistic respiration sensing using cascaded sonar on commodity devices. In *2022 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 164–171. IEEE, 2022.

[49] Edward Jay Wang, William Li, Doug Hawkins, Terry Gernsheimer, Colette Norby-Slycord, and Shwetak N Patel. Hemaapp: noninvasive blood screening of hemoglobin using smartphone cameras. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 593–604, 2016.

[50] Edward Jay Wang, Junyi Zhu, Mohit Jain, Tien-Jui Lee, Elliot Saba, Lama Nachman, and Shwetak N Patel. Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera. In *Proceedings of the 2018 CHI conference on human factors in computing Systems*, pages 1–9, 2018.

[51] Chunjong Park, Hung Ngo, Libby Rose Lavitt, Vincent Karuri, Shiven Bhatt, Peter Lubell-Doughtie, Anuraj H Shankar, Leonard Ndwiga, Victor Osoti, Juliana K Wambua, et al. The design and evaluation of a mobile system for rapid diagnostic test interpretation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1):1–26, 2021.

[52] Anandghan Waghmare, Farshid Salemi Parizi, Jason Hoffman, Yuntao Wang, Matthew Thompson, and Shwetak Patel. Glucoscreen: A smartphone-based readerless glucose test strip for prediabetes screening. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(1):1–20, 2023.

[53] Alex Mariakakis, Edward Wang, Shwetak Patel, and Mayank Goel. Challenges in realizing smartphone-based health sensing. *IEEE Pervasive Computing*, 18(2):76–84, apr 2019.

[54] Kyra H Grantz, Hannah R Meredith, Derek AT Cummings, C Jessica E Metcalf, Bryan T Grenfell, John R Giles, Shruti Mehta, Sunil Solomon, Alain Labrique, Nishant Kishore, et al. The use of mobile phone data to inform analysis of covid-19 pandemic epidemiology. *Nature communications*, 11(1):1–8, 2020.

[55] Robert P Hirten, Matteo Danieletto, Lewis Tomalin, Katie Hyewon Choi, Micol Zweig, Eddye Golden, Sparshdeep Kaur, Drew Helmus, Anthony Biello, Renata Pyzik, et al. Use of physiological data from a wearable device to identify sars-cov-2 infection and symptoms and predict covid-19 diagnosis: observational study. *Journal of medical Internet research*, 23(2):e26107, 2021.

[56] Mike A Merrill and Tim Althoff. Self-supervised pretraining and transfer learning enable flu and covid-19 predictions in small mobile sensing datasets. *arXiv preprint arXiv:2205.13607*, 2022.

[57] Srinivasan Venkatramanan, Adam Sadilek, Arindam Fadikar, Christopher L Barrett, Matthew Biggerstaff, Jiangzhuo Chen, Xerxes Dotiwalla, Paul Eastham, Bryant Gipson, Dave Higdon, et al. Forecasting influenza activity using machine-learned mobility map. *Nature communications*, 12(1):1–12, 2021.

[58] Daniel J Niven, Jonathan E Gaudet, Kevin B Laupland, Kelly J Mrklas, Derek J Roberts, and Henry Thomas Stelfox. Accuracy of peripheral thermometers for estimating temperature: a systematic review and meta-analysis. *Annals of internal medicine*, 163(10):768–777, 2015.

[59] Sarah Farnell, Lorraine Maxwell, Seok Tan, Andrew Rhodes, and Barbara Philips. Temperature measurement: Comparison of non-invasive methods used in adult critical care. *Journal of Clinical Nursing*, 14(5):632–639, 2005.

[60] Lari Lawson, Elizabeth J Bridges, Isabelle Ballou, Ruthe Eraker, Sheryl Greco, Janie Shively, and Vanessa Sochulak. Accuracy and precision of noninvasive temperature measurement in adult intensive care patients. *American journal of critical care*, 16(5):485–496, 2007.

[61] Valentina Pecoraro, Davide Petri, Giorgio Costantino, Alessandro Squizzato, Lorenzo Moja, Gianni Virgili, and Ersilia Lucenteforte. The diagnostic accuracy of digital, infrared

and mercury-in-glass thermometers in measuring body temperature: a systematic review and network meta-analysis. *Internal and emergency medicine*, 16(4):1071–1083, 2021.

[62] Fan Lai, Xin Li, Qi Wang, Yingjuan Luo, Xin Wang, Xiuhua Huang, Jiajia Zhang, Jieru Peng, Qin Wang, Li Fan, et al. Reliability of non-contact infrared thermometers for fever screening under covid-19. *Risk management and healthcare policy*, 15:447, 2022.

[63] Minh Long Hoang, Marco Carratù, Vincenzo Paciello, and Antonio Pietrosanto. Body temperature—indoor condition monitor and activity recognition by mems accelerometer based on iot-alert system for people in quarantine due to covid-19. *Sensors*, 21(7):2313, 2021.

[64] Diego Barrettino, Thomas Gisler, Christoph Zumbühl, Christian Di Battista, Raphael Kummer, and Markus Thalmann. Wearable medical device for remote monitoring the health of elderly people at home. In *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6. IEEE, 2020.

[65] Sujit Shinde, Swapna Agarwal, Dibyanshu Jaiswal, Avik Ghose, Sanjay Kimbahune, and Pravin Pillai. Thermotrak: smartphone based real-time fever screening: demo abstract. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 639–640, 2020.

[66] Peter Wei, Chenye Yang, and Xiaofan Jiang. Low-cost multi-person continuous skin temperature sensing system for fever detection. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 705–706, 2020.

[67] Huawei's clever new smartphone can take your temperature. `https://www.fastcompany.com/90513189/huaweis-new-smartphone-can-take-your-temperature`. Accessed: 2021-02-18.

[68] Amee Trivedi, Phuthipong Bovornkeeratiroj, Joseph Breda, Prashant Shenoy, Jay Taneja, and David Irwin. Phone-based ambient temperature sensing using opportunistic crowdsensing and machine learning. *Sustainable Computing: Informatics and Systems*, 29:100479.

[69] Aart Overeem, JC R. Robinson, Hidde Leijnse, Gert-Jan Steeneveld, BK P. Horn, and Remko Uijlenhoet. Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters*, 40(15):4081–4085, 2013.

[70] Joseph Breda, Amee Trivedi, Chulabhaya Wijesundara, Phuthipong Bovornkeeratiroj, David Irwin, Prashant Shenoy, and Jay Taneja. Hot or not: Leveraging mobile devices for ubiquitous temperature sensing. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pages 41–50, 2019.

[71] Liang He, Youngmoon Lee, and Kang G Shin. Mobile device batteries as thermometers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–21, 2020.

[72] Nguyen Hai Chau. Estimation of air temperature using smartphones in different contexts. *Journal of Information and Telecommunication*, 3(4):494–507, 2019.

[73] Joseph Breda and Shwetak Patel. Intuitive and ubiquitous fever monitoring using smartphones and smartwatches. *arXiv preprint arXiv:2106.11855*, 2021.

[74] Sanghoon Jun, Kilho Lee, and Jinkyu Lee. Thermobile: Measuring body temperature using a mobile device. *IEEE Sensors Journal*, 2022.

[75] National Highway Traffic Safety Administration et al. Fatality analysis reporting system, 2011.

[76] Federal Highway Administration. Highway safety information system.

[77] Lana Samara, Paul St-Aubin, Franz Loewenherz, Noah Budnick, and Luis Miranda-Moreno. Video-based network-wide surrogate safety analysis to support a proactive network screening using connected cameras: Case study in the city of bellevue (wa) united states. In *Proceedings of the Transportation Research Board 100th Annual Meeting, Washington, DC, USA*, pages 9–13, 2020.

[78] Upride. `https://upride.cc/`. Accessed: 1-24-2024.

[79] Ahmet-Serdar Karakaya, Jonathan Hasenburg, and David Bermbach. Simra: Using crowdsourcing to identify near miss hotspots in bicycle traffic. *Pervasive and Mobile Computing*, 67:101197, 2020.

[80] Apple. Use crash detection on iphone or apple watch to call for help in an accident. https://support.apple.com/en-us/104959. Accessed: 1-20-2024.

[81] The Japan News. Japan skiers' smartphones making unnecessary emergency calls. https://japannews.yomiuri.co.jp/society/general-news/20230129-87465/. Accessed: 1-20-2024.

[82] Richard Manton, Henrike Rau, Frances Fahy, Jerome Sheahan, and Eoghan Clifford. Using mental mapping to unpack perceived cycling risk. *Accident Analysis & Prevention*, 88:138–149, 2016.

[83] Kotaro Hara, Victoria Le, and Jon E. Froehlich. Combining crowdsourcing and google street view to identify street-level accessibility problems. In *Proceedings of the 31st international conference on Human factors in computing systems*, CHI '13, 2013.

[84] Manaswi Saha, Mikey Saugstad, Teja Maddali, Aileen Zeng, Ryan Holland, Steven Bower, Aditya Dash, Sage Chen, Anthony Li, Kotaro Hara, and Jon E. Froehlich. Project sidewalk: A web-based crowdsourcing tool for collecting sidewalk accessibility data at scale. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.

[85] Chu Li, Katrina Ma, Michael Saugstad, Kie Fujii, Molly Delaney, Yochai Eisenberg, Delphine Labbé, Judy L Shanley, Devon Snyder, Florian P Thomas, et al. "i never realized sidewalks were a big deal": A case study of a community-driven sidewalk audit using project sidewalk. 2024.

[86] Fabio Miranda, Maryam Hosseini, Marcos Lage, Harish Doraiswamy, Graham Dove, and Cláudio T Silva. Urban mosaic: Visual exploration of streetscapes using large-scale image data. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2020.

[87] Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 116–125, 2014.

[88] Richard A Plunz, Yijia Zhou, Maria Isabel Carrasco Vintimilla, Kathleen Mckeown, Tao Yu, Laura Uguccioni, and Maria Paola Sutto. Twitter sentiment in new york city parks as measure of well-being. *Landscape and urban planning*, 189:235–246, 2019.

[89] Simone Mora, Amin Anjomshoaa, Tom Benson, Fábio Duarte, and Carlo Ratti. Towards large-scale drive-by sensing with multi-purpose city scanner nodes. In *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pages 743–748. IEEE, 2019.

[90] Joshua S Apte and Chirag Manchanda. High-resolution urban air pollution mapping. *Science*, 385(6707):380–385, 2024.

[91] Joshua S Apte, Kyle P Messier, Shahzad Gani, Michael Brauer, Thomas W Kirchstetter, Melissa M Lunden, Julian D Marshall, Christopher J Portier, Roel CH Vermeulen, and Steven P Hamburg. High-resolution air pollution mapping with google street view cars: exploiting big data. *Environmental science & technology*, 51(12):6999–7008, 2017.

[92] Yeran Sun, Yunyan Du, Yu Wang, and Liyuan Zhuang. Examining associations of environmental characteristics with recreational cycling behaviour by street-level strava data. *International journal of environmental research and public health*, 14(6):644, 2017.

[93] Elliot Mbunge. Integrating emerging technologies into covid-19 contact tracing: Opportunities, challenges and pitfalls. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(6):1631–1636, 2020.

[94] Brian Ferris, Kari Watkins, and Alan Borning. Onebusaway: results from providing real-time arrival information for public transit. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1807–1816, 2010.

[95] Shane B Eisenman, Emiliano Miluzzo, Nicholas D Lane, Ronald A Peterson, Gahng-Seop Ahn, and Andrew T Campbell. Bikenet: A mobile sensing system for cyclist experience mapping. *ACM Transactions on Sensor Networks (TOSN)*, 6(1):1–39, 2010.

[96] Joseph Breda, Mastafa Springston, Alex Mariakakis, and Shwetak Patel. Feverphone: Accessible core-body temperature sensing for fever monitoring using commodity smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(1):1–23, 2023.

[97] Amos Grünebaum, Frank A Chervenak, Laurence B McCullough, Joachim W Dudenhausen, Eran Bornstein, and Philip A Mackowiak. How fever is defined in covid-19 publications: a disturbing lack of precision. *Journal of Perinatal Medicine*, 49(3):255–261, 2021.

[98] Wei-jie Guan, Zheng-yi Ni, Yu Hu, Wen-hua Liang, Chun-quan Ou, Jian-xing He, Lei Liu, Hong Shan, Chun-liang Lei, David SC Hui, et al. Clinical characteristics of coronavirus disease 2019 in china. *New England journal of medicine*, 382(18):1708–1720, 2020.

[99] Guy Boivin, Isabelle Hardy, Guy Tellier, and Jean Maziade. Predicting influenza infections during epidemics with use of a clinical case definition. *Clinical infectious diseases*, 31(5):1166–1169, 2000.

[100] Christine E Long, Caroline B Hall, Coleen K Cunningham, Leonard B Weiner, Kathleen P Alger, Maria Gouveia, Carol B Colella, Kenneth C Schnabel, and William H Barker. Influenza surveillance in community-dwelling elderly compared with children. *Archives of family medicine*, 6(5):459, 1997.

[101] ThM Govaert, GJ Dinant, K Aretz, and JA Knottnerus. The predictive value of influenza symptomatology in elderly people. *Family practice*, 15(1):16–22, 1998.

[102] S Kohno, T Otsubo, K Hara, Y Tomii, and J Seki. Program and abstracts of the 35th interscience conference on antimicrobial agents and chemotherapy. 1995.

[103] Fever. `https://www.cedars-sinai.org/health-library/diseases-and-conditions/f/fever.html`. Accessed: 2022-06-30.

[104] Chen Ge, Xie Jiarong, Dai Guangli, Peijun Zheng, Hu Xiaqing, Lu Hongpeng, Xu Lei, Chen Xueqin, and Chen Xiaomin. Validity of the use of wrist and forehead temperatures in screening the general population for covid-19: A prospective real-world study. *Iranian Journal of Public Health*, 49(Suppl 1):57, 2020.

[105] RC Parmar, DR Sahu, SB Bavdekar, et al. Knowledge, attitude and practices of parents of children with febrile convulsion. *Journal of postgraduate medicine*, 47(1):19, 2001.

[106] Anwar Ahmed and Birhanu Sintayehu. Implementation of covid-19 protection protocols and its implication on learning & teaching in public schools. *Heliyon*, 8(5):e09362, 2022.

[107] Forsad Al Hossain, Andrew A Lover, George A Corey, Nicholas G Reich, and Tauhidur Rahman. Flusense: a contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–28, 2020.

[108] K Taylor and L Siliver. In emerging economies, smartphone adoption has grown more quickly among younger generations. *Pew Research Cetner, Feb*, 5, 2019.

[109] Aaron C Miller, Inder Singh, Erin Koehler, and Philip M Polgreen. A smartphone-driven thermometer application for real-time population-and individual-level influenza surveillance. *Clinical Infectious Diseases*, 67(3):388–397, 2018.

[110] Michela Masè, Andreas Werner, Gabriel Putzer, Giovanni Avancini, Marika Falla, Hermann Brugger, Alessandro Micarelli, and Giacomo Strapazzon. Low ambient temperature exposition impairs the accuracy of a non-invasive heat-flux thermometer. *Frontiers in Physiology*, page 283, 2022.

[111] Frances Doyle, W John Zehner, and Thomas E Terndrup. The effect of ambient temperature extremes on tympanic and oral temperatures. *The American journal of emergency medicine*, 10(4):285–289, 1992.

[112] Charmaine Childs. Body temperature and clinical thermometry. *Handbook of clinical neurology*, 157:467–482, 2018.

[113] Paul Webb. Temperatures of skin, subcutaneous tissue, muscle and core in resting men in cold, comfortable and hot conditions. *European journal of applied physiology and occupational physiology*, 64(5):471–476, 1992.

[114] Rainer Lenhardt and Daniel I Sessler. Estimation of mean body temperature from mean skin and core temperature. *The Journal of the American Society of Anesthesiologists*, 105(6):1117–1121, 2006.

[115] Hsuan-Yu Chen, Andrew Chen, and Chiachung Chen. Investigation of the impact of infrared sensors on core body temperature monitoring by comparing measurement sites. *Sensors*, 20(10):2885, 2020.

[116] Temperature measurement for patients with fever. `https://www.uspharmacist.com/article/temperature-measurement-for-patients-with-fever`. Accessed: 2022-10-25.

[117] Yoshitaka Inui, Satoshi Hirayama, and Tadashi Tanaka. Detailed estimation method of heat generation during charge/discharge in lithium-ion battery using equivalent circuit. *Electronics and Communications in Japan*, 102(12):3–14, 2019.

[118] Thermometers: Understand the options. `https://www.mayoclinic.org/diseases-conditions/fever/in-depth/thermometers/art-20046737`. Accessed: 2022-07-31.

[119] Ricardo Morán-Navarro, Javier Courel-Ibáñez, Alejandro Martínez-Cava, Elena Conesa-Ros, Alejandro Sánchez-Pay, Ricardo Mora-Rodriguez, and Jesús G Pallarés. Validity of skin, oral and tympanic temperatures during exercise in the heat: effects of wind and sweat. *Annals of Biomedical Engineering*, 47(1):317–331, 2019.

[120] JA Kistemaker, EA Den Hartog, and HAM Daanen. Reliability of an infrared forehead skin thermometer for core temperature measurements. *Journal of medical engineering & technology*, 30(4):252–261, 2006.

[121] Temperature: Digital and glass thermometers. `https://www.nationwidechildrens.org/family-resources-education/health-wellness-and-safety-resources/helping-hands/temperature-digital-and-glass-thermometers`. Accessed: 2022-08-15.

[122] Nivedita Bijlani, Ramin Nilforooshan, Samaneh Kouchaki, et al. An unsupervised data-driven anomaly detection approach for adverse health conditions in people living with dementia: Cohort study. *JMIR aging*, 5(3):e38211, 2022.

[123] Yaguo Lei, Naipeng Li, Liang Guo, Ningbo Li, Tao Yan, and Jing Lin. Machinery health prognostics: A systematic review from data acquisition to rul prediction. *Mechanical systems and signal processing*, 104:799–834, 2018.

[124] Joon-Ho Choi and Vivian Loftness. Investigation of human body skin temperatures as a bio-signal to indicate overall thermal sensations. *Building and Environment*, 58:258–269, 2012.

[125] Centers for Disease Control and Prevention. Estimated influenza-related illnesses, medical visits, hospitalizations, and deaths in the united states — 2019–2020 influenza season, October 2021. Archived CDC Report.

[126] Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020.

[127] Sarah E Smith-Jeffcoat. Sars-cov-2 viral shedding and rapid antigen test performance—respiratory virus transmission network, november 2022–may 2023. *MMWR. Morbidity and Mortality Weekly Report*, 73, 2024.

[128] Benjamin J Smith, David J Price, Douglas Johnson, Bruce Garbutt, Michelle Thompson, Louis B Irving, Mark Putland, and Steven YC Tong. Influenza with and without fever: clinical predictors and impact on outcomes in patients requiring hospitalization. In *Open forum infectious diseases*, volume 7, page ofaa268. Oxford University Press US, 2020.

[129] Alison Han, Lindsay M Czajkowski, Amanda Donaldson, Holly Ann Baus, Susan M Reed, Rani S Athota, Tyler Bristol, Luz Angela Rosas, Adriana Cervantes-Medina, Jeffery K Taubenberger, et al. A dose-finding study of a wild-type influenza a (h3n2) virus in a healthy volunteer human challenge model. *Clinical Infectious Diseases*, 69(12):2082–2090, 2019.

[130] Alison Han, Jiat-Ling Poon, John H Powers, Nancy K Leidy, Ren Yu, and Matthew J Memoli. Using the influenza patient-reported outcome (flu-pro) diary to evaluate symptoms of influenza viral infection in a healthy human challenge model. *BMC Infectious Diseases*, 18:1–9, 2018.

[131] Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the national academy of sciences*, 102(43):15545–15550, 2005.

[132] Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P Mesirov, and Pablo Tamayo. The molecular signatures database hallmark gene set collection. *Cell systems*, 1(6):417–425, 2015.

[133] Aimee K Zaas, Minhua Chen, Jay Varkey, Timothy Veldman, Alfred O Hero, Joseph Lucas, Yongsheng Huang, Ronald Turner, Anthony Gilbert, Robert Lambkin-Williams,

et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell host & microbe*, 6(3):207–217, 2009.

[134] James F Sallis, Myron F Floyd, Daniel A Rodríguez, and Brian E Saelens. Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation*, 125(5):729–737, 2012.

[135] Christopher E Kline, Charles H Hillman, Bonny Bloodgood Sheppard, Bethany Tennant, David E Conroy, Richard F Macko, David X Marquez, Steven J Petruzzello, Kenneth E Powell, and Kirk I Erickson. Physical activity and sleep: An updated umbrella review of the 2018 physical activity guidelines advisory committee report. *Sleep medicine reviews*, 58:101489, 2021.

[136] Greig Logan, Camilla Somers, Graham Baker, Hayley Connell, Stuart Gray, Paul Kelly, Emma McIntosh, Paul Welsh, Cindy M Gray, and Jason MR Gill. Benefits, risks, barriers, and facilitators to cycling: a narrative review. *Frontiers in Sports and Active Living*, 5, 2023.

[137] Wu Chen, Trine Agervig Carstensen, Ranran Wang, Sybil Derrible, David Rojas Rueda, Mark J Nieuwenhuijsen, and Gang Liu. Historical patterns and sustainability implications of worldwide bicycle ownership and use. *Communications Earth & Environment*, 3(1):1–9, 2022.

[138] The United States Environmental Protection Agency. Inventory of u.s. greenhouse gas emissions and sinks: 1990-2021. `https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2021`, 2021. Accessed: 1-31-2024.

[139] Rui Jiang, Peng Wu, and Chengke Wu. Driving factors behind energy-related carbon emissions in the us road transport sector: A decomposition analysis. *International journal of environmental research and public health*, 19(4):2321, 2022.

[140] United Nations. Take action for the sustainable development goals. `https://www.un.org/sustainabledevelopment/sustainable-development-goals/`. Accessed: 1-21-2024.

[141] World Health Organization. Cyclist safety: an information resource for decision-makers and practitioners. `https://iris.who.int/bitstream/handle/10665/336393/9789240013698-eng.pdf?sequence=1`, 2020. Accessed: 1-21-2024.

[142] Richard Kuzmyak and Jennifer Dill. Walking and bicycling in the united states. Accessed: 1-21-2024.

[143] Alexandra Knight and Samuel G Charlton. Protected and unprotected cycle lanes' effects on cyclists' behaviour. *Accident Analysis & Prevention*, 171:106668, 2022.

[144] Ricardo Marqués and Vicente Hernández-Herrador. On the effect of networks of cycle-tracks on the risk of cycling. the case of seville. *Accident Analysis & Prevention*, 102:181–190, 2017.

[145] William Riggs. Perception of safety and cycling behaviour on varying street typologies: Opportunities for behavioural economics and design. *Transportation research procedia*, 41:204–218, 2019.

[146] Kevin Lynch. *The Image of the City.* Massachusetts Institute of Technology Press, Cambridge, Massachusetts, 1960.

[147] Peter L Jacobsen. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury prevention*, 21(4):271–275, 2015.

[148] Murilo Ribeiro Rodrigues, Antônio Nélson Rodrigues da Silva, and Inaian Pignatti Teixeira. Assessing the applicability of the cyclists' level of traffic stress (lts) classification to a medium-sized city in a developing country. *Journal of Transport & Health*, 24:101321, 2022.

[149] Anneka R Lawson, Vikram Pakrashi, Bidisha Ghosh, and WY Szeto. Perception of safety of cyclists in dublin city. *Accident Analysis & Prevention*, 50:499–511, 2013.

[150] Kirsten K Davison, Jessica L Werder, and Catherine T Lawson. Peer reviewed: Children's active commuting to school: Current knowledge and future directions. *Preventing chronic disease*, 5(3), 2008.

[151] Nicholas N Ferenchak and Wesley E Marshall. Validation of bicycle level of traffic stress and perceived safety for children. *Transportation research record*, 2674(4):397–406, 2020.

[152] Sasank Reddy, Katie Shilton, Gleb Denisov, Christian Cenizal, Deborah Estrin, and Mani Srivastava. Biketastic: sensing and mapping for better biking. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1817–1820, 2010.

[153] Oddrun Helen Hagen and Maja Karoline Rynning. Promoting cycling through urban planning and development: a qualitative assessment of bikeability. *Urban, planning and transport research*, 9(1):276–305, 2021.

[154] Susan Handy, Bert Van Wee, and Maarten Kroesen. Promoting cycling for transport: research needs and challenges. *Transport reviews*, 34(1):4–24, 2014.

[155] Colin Ferster, Trisalyn Nelson, Karen Laberee, and Meghan Winters. Mapping bicycling exposure and safety risk using strava metro. *Applied geography*, 127:102388, 2021.

[156] David Castells-Graells, Christopher Salahub, and Evangelos Pournaras. On cycling risk and discomfort: urban safety mapping and bike route recommendations. *Computing*, 102:1259–1274, 2020.

[157] Peter G Furth, Theja VVK Putta, and Paul Moser. Measuring low-stress connectivity in terms of bike-accessible jobs and potential bike-to-work trips. *Journal of transport and land use*, 11(1):815–831, 2018.

[158] Strava. Strava. `https://www.strava.com/register-now?utm_source=google&utm_medium=cpc&utm_campaign=RSR_B_SR_AllUser_US&utm_term=strava&utm_content=Brand_Universal&gad_source=1&gclid=Cj0KCQiAn-2tBhDVARIsAGmStVlXtHldyIcNdBpJ7T_pG40TD05cksPKO6aNxF5SwTN0cYFhjJcS90saAoJhEALw_wcB`. Accessed: 2-1-2024.

[159] Diego Pajarito, Suzanne Maas, Maria Attard, and Michael Gould. Mapping cyclists routes: involving citizens in collecting open cycling data. In *Proceedings of the AGILE*, page 7, 2018.

[160] University College London. 100 cyclists project. `https://www.ucl.ac.uk/civil-environmental-geomatic-engineering/research-projects/2022/oct/100-cyclists-project`. Accessed: 1-31-2024.

[161] Kevin Gildea, Daniel Hall, and Ciaran Simms. Configurations of underreported cyclist-motorised vehicle and single cyclist collisions: Analysis of a self-reported survey. *Accident Analysis & Prevention*, 159:106264, 2021.

[162] Herbert William Heinrich. *Industrial Accident Prevention. A Scientific Approach.* McGraw-Hill, 1941.

[163] Alexander Rasch, Sara Moll, Griselda López, Alfredo Garcia, and Marco Dozza. Drivers' and cyclists' safety perceptions in overtaking maneuvers. *Transportation research part F: traffic psychology and behaviour*, 84:165–176, 2022.

[164] Bundersministerium der justiz. Road traffic regulations (stvo) § 5 overtaking, 2024.

[165] Code de la Route. Royal decree of 1 december 1975 establishing general regulations on road traffic policing and the use of public roads, 1975.

[166] Tourism Shane Ross Minister for Transport and Sport. S.i. no. 495/2019 - road traffic (traffic and parking) (amendment) regulations 2019, 2019.

[167] National Conference of State Legislatures. Safely passing bicyclists chart. `https://www.ncsl.org/transportation/safely-passing-bicyclists-chart`. Accessed: 1-24-2024.

[168] SHIMPACHI YOSHIDA. Revised traffic law aimed at making roads safer for cyclists, 2023.

[169] Google. Google fit. `https://www.google.com/fit/`. Accessed: 4-11-2024.

[170] Bike Citizen. Bike citizen home page. `https://www.bikecitizens.net/en`. Accessed: 4-11-2024.

[171] Gustavo Romanillos, Martin Zaltz Austwick, Dick Ettema, and Joost De Kruijf. Big data and cycling. *Transport Reviews*, 36(1):114–133, 2016.

[172] Gian-Luca Savino, Tamara Von Sawitzky, Andrii Matviienko, Miriam Sturdee, Paweł W Woźniak, Markus Löchtefeld, Andrew L Kun, Andreas Riener, and Jonna Häkkilä. Cycling@ chi: Towards a research agenda for hci in the bike lane. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–5, 2021.

[173] Martin Porcheron, Leigh Clark, Stuart Alan Nicholson, and Matt Jones. Cyclists' use of technology while on their bike. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2023.

[174] Ammar Al-Taie, Graham Wilson, Euan Freeman, Frank Pollick, and Stephen Anthony Brewster. Light it up: Evaluating versatile autonomous vehicle-cyclist external human-machine interfaces. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2024.

[175] Andrii Matviienko, Florian Müller, Dominik Schön, Paul Seesemann, Sebastian Günther, and Max Mühlhäuser. Bikear: Understanding cyclists' crossing decision-making at uncontrolled intersections using augmented reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2022.

[176] Andrii Matviienko, Mario Boot, Andreas Löcken, Bastian Pfleging, Markus Löchtefeld, Tamara Von Sawitzky, Gian-Luca Savino, Miriam Sturdee, Josh Andres, Kristy Elizabeth Boyer, et al. Learning from cycling: Discovering lessons learned from cyclinghci. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–5, 2024.

[177] Feng Guo, Sheila G Klauer, Jonathan M Hankey, and Thomas A Dingus. Near crashes as crash surrogate for naturalistic driving studies. *Transportation Research Record*, 2147(1):66–74, 2010.

[178] Amreen A Imrit, Jaimy Fischer, Timothy CY Chan, Shoshanna Saxe, and Madeleine Bonsma-Fisher. A street-specific analysis of level of traffic stress trends in strava bicycle ridership and its implications for low-stress bicycling routes in toronto. *Findings*, 2024.

[179] Duncan Rowland, Martin Flintham, Leif Oppermann, Joe Marshall, Alan Chamberlain, Boriana Koleva, Steve Benford, and Citlali Perez. Ubikequitous computing: designing interactive experiences for cyclists. In *Proceedings of the 11th international conference on human-computer interaction with mobile devices and services*, pages 1–11, 2009.

[180] Commonwealth of Massachusetts. Massachusetts General Laws, Chapter 90, Section 14: Precautions for Safety of Other Travelers, 2024. Accessed: 2024-11-09.

[181] Andrew Fox, BVK Vijaya Kumar, Jinzhu Chen, and Fan Bai. Crowdsourcing undersampled vehicular sensor data for pothole detection. In *2015 12th annual IEEE international conference on sensing, communication, and networking (SECON)*, pages 515–523. IEEE, 2015.

[182] Domenic V Cicchetti and Sara A Sparrow. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American journal of mental deficiency*, 86(2):127–137, 1981.

[183] Raymond Scupin. The kj method: A technique for analyzing data derived from japanese ethnology. *Human organization*, 56(2):233–237, 1997.

[184] Susan C Weller, Ben Vickers, H Russell Bernard, Alyssa M Blackburn, Stephen Borgatti, Clarence C Gravlee, and Jeffrey C Johnson. Open-ended interview questions and saturation. *PloS one*, 13(6):e0198606, 2018.

[185] Anthony LaMarca, Yatin Chawathe, Sunny Consolvo, Jeffrey Hightower, Ian Smith, James Scott, Timothy Sohn, James Howard, Jeff Hughes, Fred Potter, et al. Place lab: Device positioning using radio beacons in the wild. In *Pervasive Computing: Third International Conference, PERVASIVE 2005, Munich, Germany, May 8-13, 2005. Proceedings 3*, pages 116–133. Springer, 2005.

[186] Justin Tyndall. Pedestrian deaths and large vehicles. *Economics of Transportation*, 26:100219, 2021.

[187] Muin J Khoury, Marta Gwinn, and John PA Ioannidis. The emergence of translational epidemiology: from scientific discovery to population health impact. *American journal of epidemiology*, 172(5):517–524, 2010.

[188] Citi Bike NYC. Citi bike: Nyc's official bike sharing system, 2025. Accessed: 2025-05-15.

[189] Lime Micromobility. Lime: E-scooter & e-bike sharing, 2025. Accessed: 2025-05-15.

[190] Joseph Breda, Daniel Campos Zamora, Shwetak Patel, and Jon E Froehlich. Nightlight: Passively mapping nighttime sidewalk light data for improved pedestrian routing. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2025.

[191] Jane Jacobs. *The Death and Life of Great American Cities.* Random House, New York, 1961.

[192] Yen-Chia Hsu, Ting-Hao Kenneth Huang, Ting-Yao Hu, Paul Dille, Sean Prendi, Ryan Hoffman, Anastasia Tsuhlares, Jessica Pachuta, Randy Sargent, and Illah Nourbakhsh. Project rise: recognizing industrial smoke emissions. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 14813–14821, 2021.

[193] Laurence Barry and Arthur Charpentier. Personalization as a promise: Can big data change the practice of insurance? *Big Data & Society*, 7(1):2053951720935143, 2020.

[194] Edward Jay Wang. *Mobile Health Sensing for Widespread Disease Screening and Chronic Management.* PhD thesis, University of Washington, Seattle, WA, 2019. Advisor: Shwetak Patel.

[195] Joseph Breda, Fadi Yousif, Beszel Hawkins, Marinela Cotoi, Miao Liu, Ray Luo, Po-Hsuan Cameron Chen, Mike Schaekermann, Samuel Schmidgall, Xin Liu, et al. Symptomai: Toward a conversational ai agent for everyday symptom assessment. *arXiv preprint arXiv:2605.04012*, 2026.

[196] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. Discrimination through optimization: How facebook's ad delivery can lead to biased outcomes. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–30, 2019.

[197] E-ZPass Interagency Group. About e-zpass, 2025. Accessed: 2025-05-15.

[198] Joseph Breda, Keyu Chen, Thomas Ploetz, and Shwetak Patel. Proxicycle: Passively mapping cyclist safety using smart handlebars for near-miss detection. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2025.